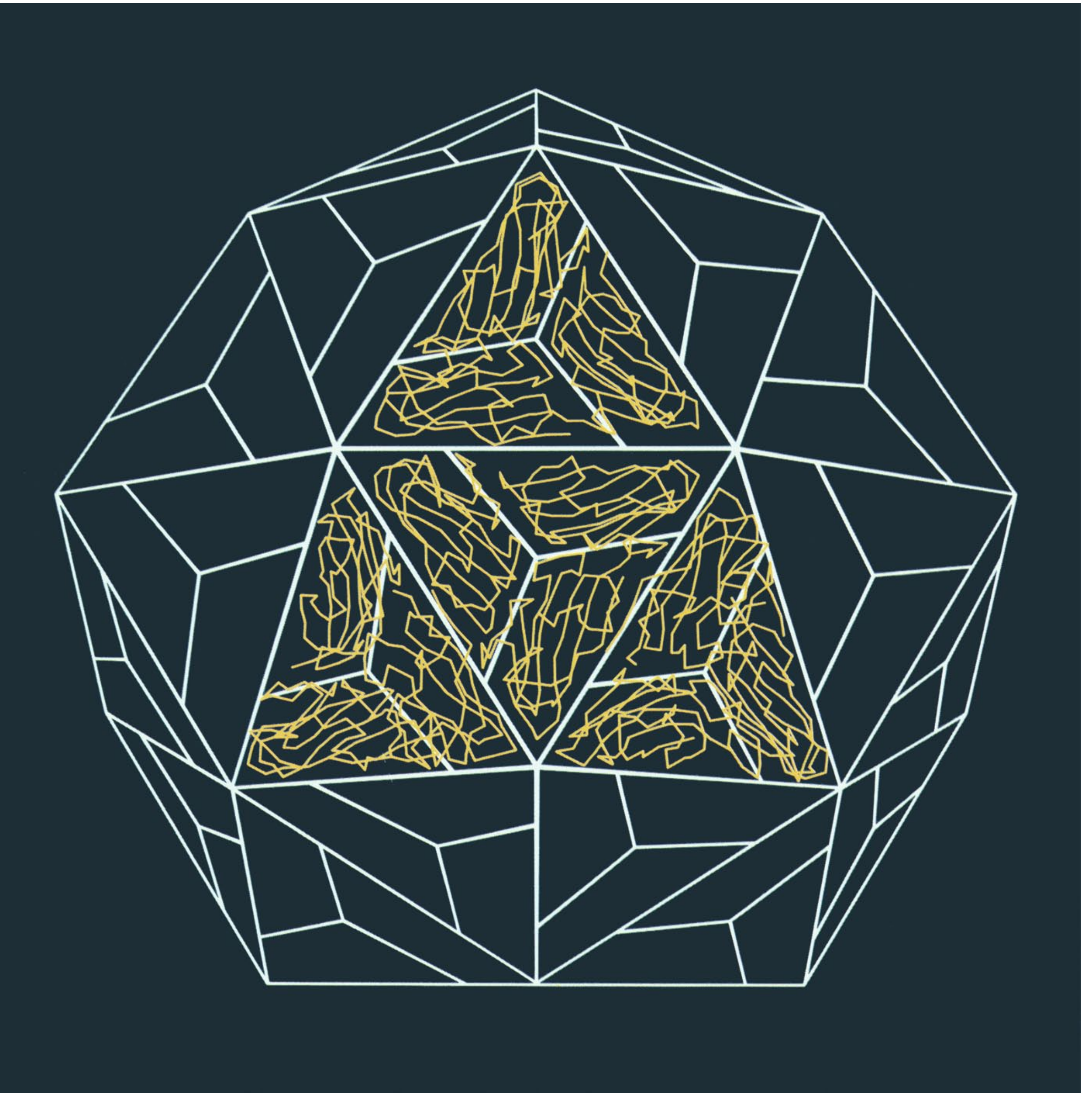
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



VACUNAS SINTETICAS

Abril 1983

350 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 79
Abril 1983

INVESTIGACION Y
CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

Colaboradores de este número:

Asesoramiento y traducción:

Ramón Pascual: *Teoría reticular del confinamiento de los quarks*; L. Enjuanes, B. G. Barreno, I. Casal, A. Sanz y C. Vela: *Vacunas sintéticas*; Javier Santamaría Salazar: *Procesos visuales ocultos*; Manuel Puigcerver: *La corona solar*; Raimundo Díaz de la Iglesia: *Computador óptico*; Joandomènec Ros: *Ranas venenosas*; Teodoro Falcón Márquez y Alfredo J. Morales: *Bóvedas de abanico*; Luis Bou: *Temas metamágicos*; J. Vilardell: *Taller y laboratorio.*

Libros:

Luis Cornudella, Pedro Pérez Puchal y José Alsina



LA PORTADA

La ilustración de la portada muestra, en azul, la estructura poliédrica de la cápside, o envoltura proteica, del virus "achaparrado peludo del tomate". La cápside consta de 180 copias de una misma proteína; en cada una de las 60 caras triangulares de la envoltura hay tres copias; 12 copias del dominio superficial de la proteína se muestran en naranja. La ilustración se basa en una representación gráfica por ordenador realizada por Arthur J. Olson; se trata de una etapa del diseño de una vacuna sintética (véase "Vacunas sintéticas", por Richard A. Lerner, en este mismo número). Mientras que una vacuna convencional dirigida contra la infección vírica contiene virus enteros, la vacuna sintética consta de un solo péptido (fragmento de una cadena proteica), sintetizado en el laboratorio, que imita parte de una proteína de la superficie del virus. Un modo de asegurarse que el péptido remede el sitio expuesto es aplicar los métodos de representación gráfica por ordenador desarrollados por T. J. O'Donnell, de Abbott, y Michael L. Connolly y Olson, del Instituto de Investigaciones de la Clínica Scripps.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 37

Condiciones de suscripción:
España:
Un año (12 números): 3.850 pesetas
Extranjero:
Un año (12 números): 52 U.S. $
Ejemplar atrasado ordinario:
350 pesetas
Ejemplar atrasado extraordinario:
450 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49 - Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137 - Madrid-30
Tel. 430 84 81
Cataluña:
Lourdes Padrós
Manuel Girona, 61, 2.° - Barcelona-34
Teléfono 204 45 83

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Ilustración de portada de Gabor Kiss

Página	Fuente
15-22	M. Alonso, J. C. Gómez-Fernández y F. M. Goñi
23	J. C. Gómez-Fernández y F. M. Goñi
25-36	Ilil Arbel
39	Arthur J. Olson, Instituto de Investigación de la Clínica Scripps
40-41	Patricia J. Wynne
42-45	Arthur J. Olson, Instituto de Investigación de la Clínica Scripps
46	Patricia J. Wynne
52	R. F. Bonifield, cortesía de la Colección Granger
54-60	George V. Kelvin
63	Observatorio de Gran Altitud
64	Jay M. Pasachoff, Williams College (*arriba*); Observatorio de Gran Altitud (*abajo*)
65	Observatorio de Gran Altitud
66	American Science & Engineering, Inc.
67	Centro de Vuelo Espacial Goddard (*arriba*), Observatorio de Gran Altitud (*abajo*)
68	Laboratorio de Investigación Naval
69-72	Allen Beechel
75-83	Walken Graphics
85	David M. Dennis
86	Patricia J. Wynne
87	Charles W. Myers, Museo Americano de Historia Natural
88	Lois W. Tice, Instituto Nacional de Artritis, Diabetes y Enfermedades Digestivas y Renales
90	Patricia J. Wynne
92	Enid Kotschnig
94	John Moss, Black Star
96-99	Alan D. Iselin
100	Walter C. Leedy, Jr., Universidad estatal de Cleveland (*arriba*); Alan D. Iselin (*abajo*)
102	Alan D. Iselin
104	John Moss, Black Star
116-121	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Pedro IV, 160 - Barcelona-5
Fotocromos reproducidos por GINSA, S.A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Napoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Cartas

Señor Director:

En su artículo sobre "Dióxido de carbono y clima mundial", INVESTIGACIÓN Y CIENCIA del mes de octubre pasado, Roger Revelle hace lo que parece ser una puesta a punto bastante ponderada del problema del CO_2 y el clima, pero se calla que la mayoría de las predicciones de los modelos de ordenador usados para estudiar el problema contradicen la realidad. Por ejemplo, al citar el trabajo de modelado de James E. Hansen y sus colaboradores, indica que el máximo calentamiento pronosticado en virtud del aumento de CO_2 se producirá en las latitudes septentrionales. Sin embargo, los propios datos de Hansen para esta región (*Science*, vol. 213, n.º 4511, págs. 957-966; 28 de agosto 1981) muestran que la temperatura media en las latitudes septentrionales ha estado disminuyendo en promedio a razón de más de 0,1 grados Celsius por decenio desde 1935, período para el que los modelos pronosticaban un aumento. Y no hay señal alguna de cambio reciente en esta tendencia. Antes bien, las observaciones mediante satélites de las variaciones en la cobertura estacional de nieve en el Hemisferio Norte presentadas por Kenneth F. Dewey y Richard R. Heim, Jr., de la Universidad de Nebraska en Lincoln, nos hablan de un incremento general del área cubierta por la nieve desde 1966 hasta 1980...

La razón primordial del ligero aumento de la temperatura global, en años recientes, estriba en que las latitudes meridionales parecen mostrar una tendencia al calentamiento. Sin embargo, como observa Revelle en relación con el trabajo de Hermann Flohn, nos hallamos ante una región del globo donde los efectos del CO_2 resultarían mínimos: de nuevo, justamente lo contrario de lo que las mediciones reales dicen...

En suma, Revelle presenta la historia típica de un grupo, muy apretado, de modelistas del clima que proclama graves consecuencias para la Tierra si se permite que el CO_2 atmosférico continúe aumentando en la misma proporción que en la actualidad. Las observaciones citadas y otros estudios empíricos indican, por contra, que la tendencia a un clima más cálido predicha por esos modelos se ha sobreestimado nada menos que en un orden de magnitud (esto es, en 10 veces). A despecho de los intentos de ese grupo y de sus representantes en la Academia Nacional de Ciencias de los Estados Unidos para negarla, la historia opuesta, contada por los datos del mundo real, no se puede ocultar por mucho más tiempo.

SHERWOOD B. IDSO

Institute for Biospheric
Research, Inc.
Tempe, Arizona

Señor Director:

El excelente artículo de Roger Revelle ofrece la paradoja de ser, a un tiempo, oportuno e inoportuno. Oportuno porque el tema del CO_2 preocupa cada vez más a los científicos, e inoportuno porque, desde que escribió su artículo, la gran erupción de El Chichón en marzo y abril ha penetrado la tropopausa e inyectado una enorme cantidad de polvo volcánico y dióxido de azufre en la estratosfera. Los informes provisionales sobre la extensión y tamaño de esta nube deducidos de medidas con lidar revelan una densidad presumiblemente mayor que la de las erupciones del Monte Agung en 1963.

Ha subrayado correctamente la incapacidad del hombre para cuantificar con precisión los efectos de calentamiento climático del CO_2. Yo añadiría la incapacidad del hombre para cuantificar los efectos de enfriamiento de la nube de polvo y aerosol de El Chichón. Los efectos contrapuestos de estos dos sucesos –uno gradual y continuo, el otro brusco y discontinuo– son evidentes. La veleidad de la naturaleza al "inferir" con los modelos humanos de las variaciones climáticas no debe dejarse pasar sin comentario. Las nubes volcánicas de polvo y aerosoles son realimentaciones negativas que retardan el calentamiento debido al CO_2. Su efecto sólo se puede determinar después de la erupción. Y ésta es imprevisible. Las variaciones de temperatura presentadas en el artículo revelan pronunciados y bruscos descensos en 1882-85, 1914-18 y 1962-64. Fechas que corresponden, respectivamente, a las erupciones del Krakatoa, Monte Katmai y Monte Agung.

El objeto de esta carta no es tanto discrepar de Revelle cuanto señalar que cualquier efecto de calentamiento debido al CO_2 puede quedar enmascarado por el efecto de enfriamiento de la nube de aerosol de El Chichón, que se espera persista varios años. Además, la tendencia de la temperatura media anual del Hemisferio Norte ha sido indefinida, más bien que de calentamiento, en el decenio de 1970. Aunque a corto plazo parece probable un enfriamiento del Hemisferio Norte, la adición del CO_2 a la atmósfera acabará por hacerse notar. El problema merece la atención que Revelle e INVESTIGACIÓN Y CIENCIA le han dispensado.

DONALD R. WIESNET

Jefe de la División
de Ciencias de la Tierra,
Administración Nacional
Oceanográfica y Atmosférica
Washington, D. C.

Señor Director:

Me es fácil simpatizar con la postura de Sherwood Idso. Me ha tocado vivir la época en que dos importantes conflictos científicos enfrentaban, por un lado, a los expertos "entendidos" y la verdad, por otro. Pero ni las pinturas paleolíticas de Altamira ni la deriva continental son buenas analogías. Los "recientes estudios empíricos" de Idso apoyan en realidad el modelo que él critica.

En uno de esos estudios, Idso ha comparado el ciclo estacional de variaciones del flujo de radiación solar en la superficie de la Tierra con la correspondiente respuesta a corto plazo de la temperatura en superficie. De este estudio y otros similares, calcula que una variación de 1 W m^{-2} (watt por metro cuadrado) en la radiación hacia abajo en superficie daría por resultado una variación casi inmediata en la temperatura en superficie de 0,2 grados Celsius. Se ha usado el mismo proceso de cálculo para comprobar empíricamente la validez de cierto número de modelos del efecto CO_2-clima. Así, Veerabhadran Ramanathan acaba de mostrar que en los supuestos de que la superficie estuviera en equilibrio radiativo, la temperatura de la atmófera permaneciese constante y no hubiera variaciones de los flujos térmicos causados por otros procesos, la duplicación instantánea del contenido de CO_2 en la atmósfera aumentaría el flujo radiativo hacia abajo en la superficie entre 1 y 1,5 W m^{-2}, y la temperatura en superficie subiría aproximadamente 0,2 grados, en estrecha concordancia con las observaciones de Idso.

Pese a la pretensión de Idso, no hay análogos experimentales conocidos que nos descubran cómo afectaría al clima esta respuesta radiativa directa al cam-

bio en concentración de CO_2 cuando estuvieran actuando todos los procesos. Así pues, hemos de echar mano de modelos climáticos teóricos.

Se sabe que a escalas de tiempo de más allá de algunas semanas, la baja atmósfera (la troposfera) está estrechamente entrelazada con la superficie por medio de movimientos convectivos y de otras clases, así como mediante los procesos radiativos. Con un aumento del contenido de CO_2, la troposfera y la superficie se calentarían porque la radiación infrarroja quedaría aprisionada en la troposfera. De acuerdo con los cálculos realizados con modelos, duplicando el contenido de CO_2 atmosférico la variación neta del flujo radiativo hacia la superficie superaría los 4 $W\ m^{-2}$ y la temperatura de la superficie subiría en aproximadamente un grado, supuesto que la cantidad de vapor de agua en la atmósfera permaneciese constante. Pero la concentración de vapor de agua en el aire aumentará al ascender la temperatura, lo que dará por resultado mayor aprisionamiento de radiación infrarroja en la troposfera. Los cálculos con modelos indican que, con esta realimentación positiva, la temperatura aumentará en aproximadamente dos grados. Hay otras varias realimentaciones positivas posibles. Por ejemplo, si las nubes suben al aumentar las temperaturas troposféricas, de manera que la temperatura y la radiación infrarroja emitida por las cimas de las nubes permaneciesen constantes, la temperatura cerca de la superficie aumentaría. Análogamente, la fusión de la nieve y hielo causada por el calentamiento inducido por el CO_2 reduciría la reflectividad de la Tierra, dando por resultado mayor absorción de la radiación solar y, por tanto, un ulterior calentamiento.

Los cálculos de los efectos combinados del calentamiento troposférico y estas realimentaciones positivas ofrecen un aumento global de temperatura estimado de $3 \pm 1{,}5$ grados al duplicarse la cantidad de CO_2, lo que representa una multiplicación por más de diez de la relativa influencia sobre la temperatura en superficie que ejercería el flujo radiativo directo, calculada por Ramanathan (y también por Idso) a partir de consideraciones de transporte radiativo en situaciones en que no hay equilibrio.

Idso afirma, como yo lo había hecho en mi artículo, que las variaciones de temperatura atmosférica registradas a lo largo de los últimos 100 años no confirman los estudios con modelos. Expuse entonces la estrecha correlación obtenida por James E. Hansen y sus colaboradores entre esas temperaturas, cuando se promediaban los datos sobre una base global, y las variaciones de temperatura calculadas a partir de un modelo que tomaba en consideración el aumento de CO_2, la fluctuación en los aerosoles volcánicos, la inercia térmica oceánica y una pequeña variación en la luminosidad del Sol. Para Michael C. MacCraken hay varios puntos débiles en esta correlación, especialmente la discrepancia entre los cálculos del modelo y la distribución según la latitud de las variaciones de temperatura observadas. Los modelos basados en equilibrio predicen una amplificación del calentamiento inducido por el CO_2 a latitudes septentrionales altas, pero las observaciones no muestran esto. En lugar de ello, indican un enfriamiento medio, en la zona de latitudes entre 23,6 grados Norte y 90 grados Norte, de aproximadamente 0,5 grados desde 1935 a 1970, seguido de un calentamiento de 0,2 grados después de 1970. La afirmación de Idso de que el aumento observado de la superficie cubierta por nieve en 1966-80 revela un enfriamiento continuo es probablemente incorrecta. En regiones muy frías, un ligero calentamiento conducirá a mayor caída de nieve porque el aire es capaz de retener y luego desprender mayor cantidad de vapor de agua.

Varios factores climáticos –la extensión de la cobertura nubosa, la respuesta del ciclo hidrológico y el comportamiento de los océanos– quizá no reciban una atención satisfactoria en los actuales modelos. Debemos concluir que hasta que se contemple una obvia tendencia al calentamiento que exceda el nivel de ruido de las fluctuaciones climáticas naturales, existirán dudas razonables y diversidad de opiniones acerca de la amplitud de los efectos climáticos del aumento de CO_2 atmosférico. Si los modelos son correctos, tal señal debería detectarse dentro de 10 o 15 años.

En cuanto a la carta de Donald R. Wiesnet, estoy de acuerdo con él excepto en un punto. Como antes he señalado, las temperaturas medias al Norte de 23,6 grados Norte subieron unos 0,2 grados Celsius entre 1970 y 1980. Sin embargo, como el propio Wiesnet insinúa, quizá fue mero "ruido" climático. Adviértase que 1981 fue el año más cálido registrado en el Hemisferio Norte.

ROGER REVELLE

Profesor de Ciencia
y Normas de Conducta Públicas,
Universidad de California en
San Diego. La Jolla

Los autores

JUAN C. GOMEZ-FERNANDEZ y FELIX M. GOÑI ("Fluidez de las membranas celulares") enseñan bioquímica en la Universidad de Murcia, el primero, y en la del País Vasco, el segundo. Ambos cursaron estudios en la Universidad de Navarra. Gómez-Fernández se doctoró en ciencias biológicas en 1975 y es también licenciado en farmacia. Realizó investigación postdoctoral, durante tres años, en la Universidad de Oxford, con G. K. Radda, y en la de Londres, con D. Chapman. Actualmente dirige un grupo de investigación sobre biomembranas, estando especialmente interesado en interacciones lípido-proteína y en enzimas de biomembranas. Goñi se doctoró en medicina en 1975 y realizó una estancia postdoctoral en la Universidad de Londres con D. Chapman. En la actualidad dirige un grupo de investigación sobre biomembranas.

CLAUDIO REBBI ("Teoría reticular del confinamiento de los quarks") es un teórico de partículas elementales, adscrito al Brookhaven National Laboratory. Nació en Trieste y estudió física en la Universidad de Turín, por la que se recibió de doctor en 1967. Realizó su trabajo postdoctoral en el Instituto de Tecnología de California. Enseñó en la Universidad de Trieste e investigó en la Organización Europea para la Investigación Nuclear (CERN), en Ginebra. En 1974 retornó a los Estados Unidos contratado por el Instituto de Tecnología de Massachusetts. Pasó a Brookhaven en 1977. Durante el año académico 1980-81 trabajó de nuevo en el CERN. Es éste el segundo artículo de Rebbi que aparece en INVESTIGACIÓN Y CIENCIA; el primero, "Solitones", se publicó en abril de 1979.

RICHARD A. LERNER ("Vacunas sintéticas") dirige el departamento de biología molecular del Instituto de Investigaciones de la Clínica Scripps de La Jolla, California. En 1964 se doctoró por la Facultad de Medicina de la Universidad de Stanford, donde permaneció luego un año como médico interno residente. Desde entonces ha trabajado para la Clínica Scripps, salvo dos años que pasó en el Instituto Wistar de Anatomía y Biología, de Filadelfia, de 1968 a 1970. Es un empedernido escritor de novelas de ciencia-ficción.

JEREMY M. WOLFE ("Procesos visuales ocultos") tiene a su cargo el curso introductorio a la psicología en el Instituto de Tecnología de Massachusetts (MIT), en cuyo centro de ciencias cognoscitivas investiga. Nacido en Londres, se licenció con premio extraordinario por la Universidad de Princeton en 1977. Realizó su trabajo de doctorado sobre visión binocular con Richard Held, en el MIT, obteniendo el título en 1981. "En los años inmediatos a mi ingreso en la universidad pensaba hacerme abogado o rabino, pero me convertí a la psicología experimental durante un trabajo de verano en el centro de investigación de los Laboratorios Bell. Pasé el tiempo sentado en una cámara oscura mirando destellos de colores apenas visibles y decidí hacer lo mismo a otras personas."

RICHARD WOLFSON ("La corona solar") enseña física en el Middlebury College. Cursó el primer ciclo universitario en el Instituto de Tecnología de Massachusetts y en el Swarthmore College, por el que se tituló en filosofía y en física en 1969. Tras un breve período de investigación en física médica en el Hospital General de Massachusetts, cursó estudios en la Universidad de Michigan. Posteriormente marchó al Dartmouth College, por el que se recibió de doctor en física en 1976. Su investigación, nos cuenta él mismo, "se centra en el comportamiento de gases y plasmas en situaciones astrofísicas, y comprende estudios de corrientes de depósito en fuentes de rayos X, así como vientos solares y estelares y la corona solar. En 1980-81 estuve contratado por el Observatorio de Gran Altitud del Centro Nacional de Investigación Atmosférica, donde llevé a cabo estudios teóricos de dinámica coronal. Me es grato agradecer el apoyo del personal del Observatorio de Gran Altitud y de muchos otros que proporcionaron las fotografías".

EITAN ABRAHAM, COLIN T. SEATON y S. DESMOND SMITH ("Computador óptico") trabajan juntos en la Universidad Heriot-Watt de Edimburgo. Abraham es investigador asociado al departamento de física, donde enseña estadística cuántica, termodinámica y física de computación. Nacido en Israel, se licenció por la Universidad de Buenos Aires en 1975. En 1979 se doctoró por la Universidad de Manchester. Seaton se licenció en física aplicada en 1979 por la de Heriot-Watt, donde en la actualidad ultima sus trabajos de doctorado. Su investigación se centra en el estudio de los efectos ópticos no lineales de los semiconductores, las operaciones de selección óptica, así como en los dispositivos para el procesado óptico de la señal. Smith es profesor de física de la Heriot-Watt y decano de su Facultad de Ciencias. Licenciado por la Universidad de Bristol, es doctor en física por la Universidad de Reading y por la de Bristol (1965 y 1966, respectivamente). Antes de ocupar su puesto actual en Heriot-Watt (1970), trabajó en el Royal Aircraft Establishment de Farnborough, en el Imperial College de Londres y en la Universidad de Reading. Simultanea su actividad académica con la dirección de Edinburgh Instruments Ltd., empresa fabricante de sistemas láser. Smith es miembro de la Royal Society de Edimburgo, del Instituto de Física (que le concedió el premio C. V. Boys en 1976) y de la Royal Society de Londres. De 1977 a 1980 fue director de la división de electrónica cuántica de la Sociedad Europea de Física.

CHARLES W. MYERS y JOHN W. DALY ("Ranas venenosas") han trabajado juntos desde 1966, "dividiendo nuestro tiempo entre campamentos en la selva y modernos laboratorios en un estudio multidisciplinario de la biología y de la bioquímica de las ranas ponzoñosas". Myers es catedrático del departamento de herpetología del Museo Americano de Historia Natural. Cursó estudios de biología en la Universidad de Florida y en la Universidad de Illinois Meridional; interrumpió sus estudios de doctorado para pasar tres años (1964-67) como científico visitante en el Laboratorio Gorgas Memorial de Panamá. Se incorporó al equipo del Museo Americano en 1968, doctorándose en ciencias por la Universidad de Kansas en 1970. Daly dirige el laboratorio de química biorgánica del Instituto Nacional de Artritis, Diabetes y Enfermedades Digestivas y Renales. Estudió en el Oregon College y se doctoró en química por la Universidad de Stanford.

WALTER C. LEEDY, JR. ("Bóvedas de abanico") es historiador de la arquitectura y medievalista adscrito a la Universidad estatal de Cleveland. Se especializó en arquitectura en la Universidad de Michigan, pasando un año en Polonia, en la Fundación Kościuszko, entre 1966 y 1967. Se doctoró en historia del arte en 1972, por el Courtauld Institute de la Universidad de Londres.

Hace...

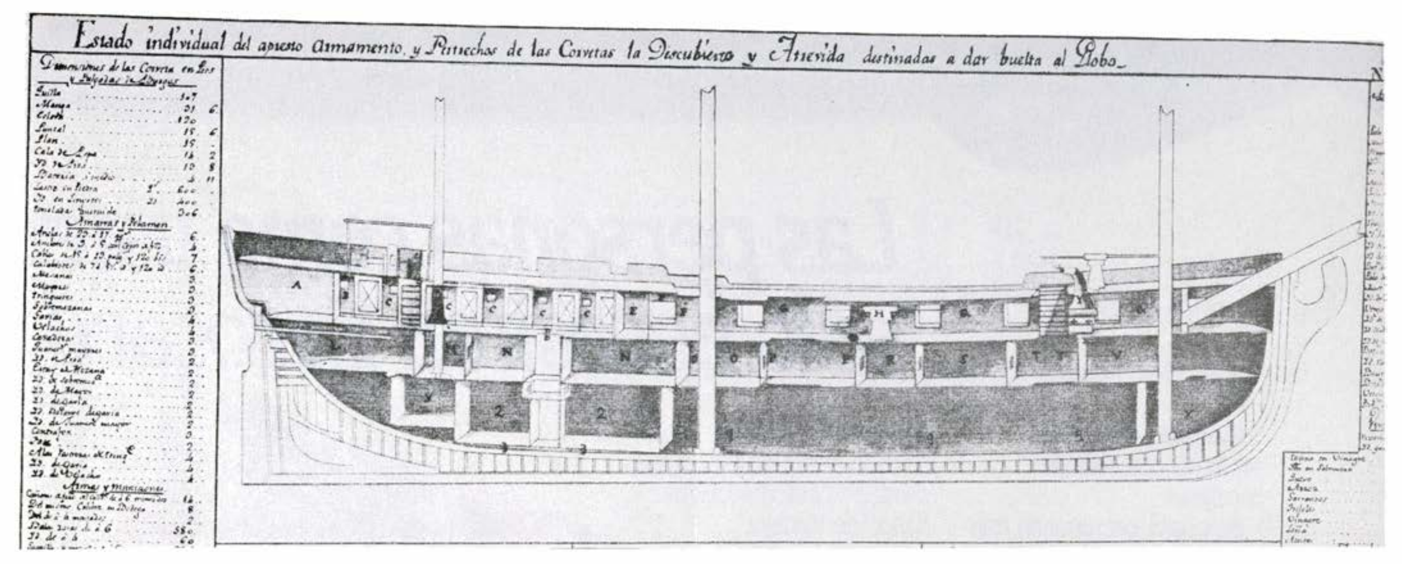
José M.ª López Piñero

... doscientos años

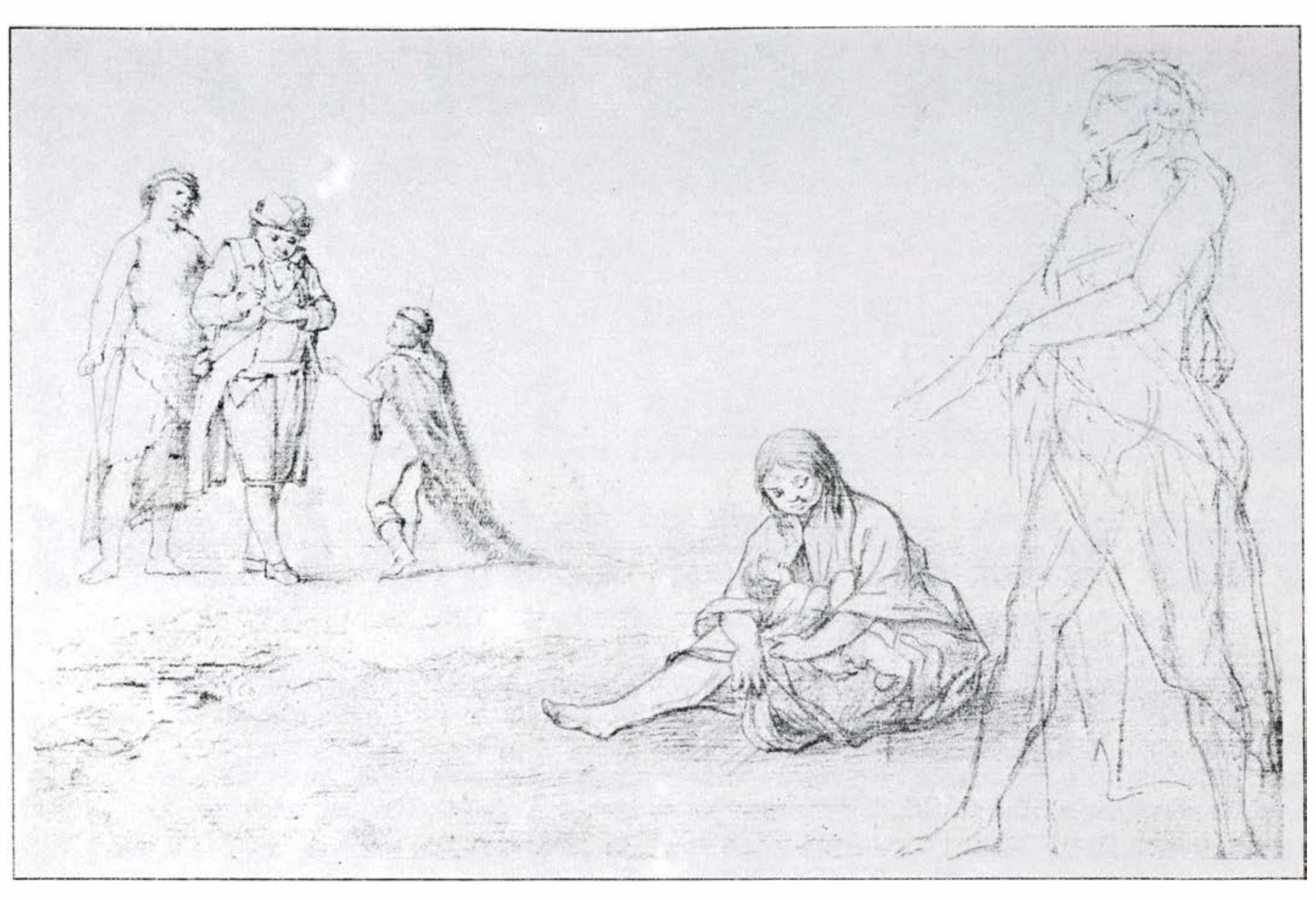
Alejandro Malaspina, recién promovido a capitán de fragata, efectuó un viaje en torno al mundo a bordo de la fragata *Astuca*. Había nacido en 1754 en la localidad italiana de Mulazzo, en el seno de una importante familia noble y, a partir de los veinte años, consagró su vida al servicio de la armada española. En 1774 sentó plaza de guardiamarina en Cádiz y dos años más tarde ascendió a alférez de fragata, realizando entonces sus primeros viajes por el Atlántico y el Pacífico. Durante el lustro siguiente participó en varias batallas navales contra la escuadra inglesa, llegando en una ocasión a caer prisionero y ser llevado a Gibraltar.

En 1789, tras ser ascendido a capitán de navío, fue puesto al frente de una ambiciosa expedición científica, destinada fundamentalmente a estudiar desde todos los puntos de vista la costa occidental del continente americano y una extensa zona del Océano Pacífico. Como todos los grandes viajes de este tipo, tenía también una finalidad política, en especial recoger información acerca del estado de las colonias americanas y de la postura de su población ante la metrópoli, así como sobre la penetración de potencias extranjeras principalmente Inglaterra y Rusia. Con destino a la expedición se construyeron *ex profeso* dos corbetas, la *Descubierta* y la *Atrevida*, la primera de las cuales se puso al mando del propio Malaspina y la segunda, al de José Bustamante Guerra, otro capitán de navío. Contaban con biblioteca y laboratorio e iban muy bien equipadas con instrumentos astronómicos y náuticos, geodésicos, meteorológicos, físicos, químicos y biológicos. Malaspina preparó cuidadosamente la expedición, consultando con numerosas personalidades científicas tanto extranjeras como españolas. Entre las primeras figuraron Joseph Jerôme de Lalande, profesor de astronomía del Collège de France, el cartógrafo del Almirantazgo británico Alexander Dalrymple y Joseph Banks, presidente de la Royal Society. De sus consejeros españoles anotaremos solamente a Gaspar de Molina, que le dio información acerca de las aplicaciones náuticas de la electricidad y sobre las nuevas máquinas para "purificar el aire", y a José de Mazarredo, quien fue, sin duda, su principal asesor. Malaspina decidió que la técnica hidrográfica que se emplearía sería la de Mazarredo, que ya había sido utilizada por Vicente Tofiño para preparar los mapas del litoral español y africano que figuran en su *Atlas marítimo de España*. Para asegurar la correcta aplicación del método, incorporó a la expedición como cartógrafo a dos discípulos de Tofiño, José Espinosa y Tello y Felipe Bauzá. Como astrónomos formaron parte del personal científico de la misma otras figuras de la marina de guerra, como Dionisio Alcalá Galiano, Juan Gutiérrez de la Concha y Juan Bernáldez. El director de los naturalistas de la expedición fue el guatemalteco Antonio Pineda, que contó con la colaboración de Luis Née y Tadeo Haenke, del botánico y cirujano Francisco Flores, del "disecador y dibujante" José Guío y del "pintor botánico y de perspectiva" José del Pozo; estos dos últimos fueron sustituidos en el curso del viaje por los dibujantes Fernando Brambila y Juan Ravent. En total, había doscientos ocho hombres a bordo de la *Descubierta* y la *Atrevida* cuando partieron de Cádiz a finales de julio de 1789.

El viaje se prolongó durante más de cinco años. El primer punto del continente americano que tocó la expedición fue Montevideo, desde donde pasó a



"Explicación de las Reparticiones del Buque: **A**) *Cámara de estudio. Contiene una pequeña biblioteca, dos mesitas y una chimenea inglesa.* **B**) *Camarote para los instrumentos astronómicos, meteorológicos y reloxes marinos.* **C**) *Cinco camarotes para oficiales de plana mayor. Igual número a estribor. El espacio comprendido entre los camarotes de una y otra banda forma una cámara con una mesa firme para 18 personas.* **D**) *Escotillón para bajar a los pañoles de pan.* **E**) *Repostería.* **F**) *Dos ranchos de oficiales de mar.* **G**) *Parte habitable para la tropa y marinería.* **H**) *Cabestrante mayor.* **Y**) *El fogón con dos destiladores.* **K**) *Comunicación del ventilador con el sollado.* **L**) *Santa Bárbara.* **M**) *Un camarote con dos catres por cada lado.* **N**) *Pañoles grandes a babor y estribor para vestuarios, efectos de cambios con los indios, harinas, repuestos de medicinas y algunos instrumentos.* **O**) *Pañoles.* **P**) *Sitios para los cables.* **Q**) *Bodega.* **R**) *Un pañol para cajas y equipajes.* **S**) *Un pañol para cargas del calafate y cirujanos y equipajes.* **T**) *Escotilla de proa.* **V**) *Despensa.* **X**) *Pañol para betunes, motonería y carbón.* **Z**) *Sucucho para efectos del condestable.* **Y**) *Pañol de pólvora.* **2**) *Pañoles de pan.* **3**) *Parte para separar el pan de la humedad por medio de un sollado."*



El naturalista Antonio Pineda, tomando notas en Patagonia, rodeado de indígenas. Dibujo a lápiz de José del Pozo (1789)

Puerto Deseado, en la Patagonia, al sur de la que hoy se llama bahía Bustamante y de las localidades que llevan los nombres de Malaspina y Bustamante. Tras una estancia en las islas Malvinas, dobló el cabo de Hornos y llegó a Concepción y Santiago de Chile. Comenzó entonces el estudio de las costas occidentales del continente americano, partiendo de las chilenas y peruanas, pasando por las de América Central y después las de México y el resto de América del Norte, hasta Alaska, en cuya bahía de Yakutat está situado el gran ventisquero que actualmente lleva el nombre de Malaspina. Desde Alaska, las corbetas volvieron a Acapulco, que sirvió de punto de partida de un viaje de año y medio por el Pacífico, cuyas principales etapas tuvieron como escenario las Filipinas –donde falleció Pineda– Nueva Zelanda, la costa oriental australiana y las islas Tonga. En julio de 1793, la expedición volvió a El Callao, donde se decidió su reparto en dos grupos: cartógrafos y naturalistas debían cruzar los Andes y llegar por tierra a la costa atlántica, mientras las corbetas con el resto del personal volvían a doblar el cabo de Hornos. Espinosa, Bauzá y Née se reunieron, en efecto, con Malaspina en Montevideo como estaba planeado, pero Haenke permaneció en América del Sur los 24 años que le quedaban de vida, sintiendo siempre ligado su trabajo a los objetivos de la expedición.

Bonifacio del Carril ha resumido de forma muy expresiva la actividad de los expedicionarios: "En todos los lugares donde se detuvieron... se ubicó astronómicamente el sitio. Se midieron y calcularon niveles. Se levantaron cartas geográficas. Se exploraron y reconocieron los alrededores. Se hicieron observaciones geológicas, botánicas y zoológicas; estudios etnográficos y lingüísticos. Se recogieron numerosas carpetas, que se fueron remitiendo a España o se conservaron en las corbetas para preparar el informe final, después del regreso". Aparte de su labor como director, la principal contribución personal de Malaspina fue de tipo cartográfico. Llegó a completar seis mapas: dos de la costa oriental de Patagonia, uno de la Tierra de Fuego y tres de la costa sudamericana del Pacífico hasta Lima; reconoció el papel destacado que Bauzá había tenido en la realización de todos ellos.

En septiembre de 1794, la *Descubierta* y la *Atrevida* volvieron a Cádiz con un material que se encuentra entre los más ricos de las expediciones científicas de la Ilustración. Malaspina presentó en la Corte los resultados científicos del viaje y también un informe político confidencial, favorable a la concesión de una amplia autonomía a las colonias americanas. Con varios de sus colaboradores, entre ellos Bauzá, se dedicó a preparar la edición de una gran obra que expusiera los frutos de la expedición. En mayo de 1795 fue ascendido a brigadier de la armada. Su influencia política era cada vez mayor, lo que alarmó a Manuel Godoy, quien lo implicó en una intriga, consiguiendo que fuera juzgado por conspiración. En abril de 1796 fue destituido de todos sus empleos y grados y condenado a diez años de prisión en el castillo de San Antón, en La Coruña. Un año más tarde, la pena de prisión le fue conmutada por la de destierro a las posesiones de su familia en Italia, donde murió. Los materiales de la expedición se dispersaron. Los cartográficos pasaron al nuevo Depósito Hidrográfico fundado en Madrid bajo la dirección de Bauzá y Espinosa, que dieron a conocer parte de ellos en una importante serie de publicaciones.

Los principales fondos procedentes de la expedición se encuentran en el Museo Naval, de Madrid ("Archivo Guillén-Expedición Malaspina, 1788-1806") y en el British Museum de Londres ("Bauzá Collection").

Fluidez de las membranas celulares

Las membranas celulares son fluidas. Del grado de fluidez dependen numerosas funciones. Los modelos de membrana semisintéticos y la aplicación de técnicas físicas permiten valorar esa propiedad y relacionarla con procesos fisiológicos

Juan C. Gómez-Fernández y Félix M. Goñi

Las células están protegidas del medio circundante por una envoltura o membrana. Quizá sea una de las propiedades menos intuitivas de ésta la de su fluidez, noción que comenzó a tomar cuerpo durante la década de los sesenta y primeros setenta con la aplicación de técnicas físicas al estudio de las biomembranas y el desarrollo de modelos de esas estructuras. Por fluidez de membrana debe entenderse la mayor o menor movilidad de que gozan las moléculas componentes de la membrana (lípidos y proteínas intrínsecas) para difundirse por la bicapa.

¿Qué factores modulan la fluidez? ¿Cómo observar tales cambios? ¿Qué consecuencias tienen éstos para la fisiología celular? A lo largo del artículo iremos abordando esas tres cuestiones. Se admite generalmente que las membranas biológicas contienen regiones formadas por lípidos anfipáticos, especialmente fosfolípidos, dispuestos en doble capa, con las partes más hidrofílicas orientadas hacia el exterior. Anfipático es un término de origen griego que significa doble sensibilidad y aquí denota el doble carácter, hidrofóbico e hidrofílico, de estas moléculas.

Las sustancias anfipáticas se disponen en agua, adoptando una conformación micelar. (Las micelas son agregados más o menos globulares en los que las partes hidrofóbicas quedan en el interior y las hidrofílicas en la superficie interaccionando con el disolvente.) Tal es el caso de las suspensiones acuosas de detergentes en circunstancias adecuadas. Los lípidos que forman parte de las membranas como sustancias anfipáticas adoptan también conformaciones micelares, pero su tendencia predominante es, con mucho, a formar películas de doble capa; así son las que vemos en las membranas biológicas.

En condiciones fisiológicas, el equilibrio entre los lípidos en forma libre, en forma micelar y en forma de bicapa está totalmente desplazado hacia esta última disposición. Cuando se suspenden, en un medio sobrante de agua, lípidos extraídos de una membrana biológica, e incluso algunos fosfolípidos puros (fosfatidilcolina), adoptan de inmediato la disposición en doble capa. Se originan así estructuras en forma de cebolla, con multitud de bicapas concéntricas, separadas por espacios acuosos. Se han empleado estas estructuras, denominadas liposomas, para diseñar modelos semisintéticos de membrana. Los liposomas multilamelares pueden convertirse, por acción de los ultrasonidos, en múltiples liposomas pequeños, de una sola bicapa. No faltan procedimientos para la obtención directa de grandes liposomas unilamelares.

A la hora de ilustrar la estructura de las biomembranas se dibuja una bicapa lipídica, con la que están asociadas las proteínas intrínsecas o extrínsecas, unidas más o menos fuertemente a la matriz lipídica.

Hallamos también asociaciones lípido-proteína similares a las de las membranas celulares en estructuras completamente distintas, las lipoproteínas del suero por ejemplo. Y han servido, a su vez, de modelos para el estudio de la interacción lípido-proteína y fluidez en membranas celulares. Pero esa línea de investigación, y en general todo lo referente al papel de las proteínas en las biomembranas, se ha llevado a cabo sobre todo en proteoliposomas, es decir, estructuras semisintéticas, parecidas a los liposomas, que contienen además una proteína o mezcla de ellas. Los proteoliposomas se preparan a veces de manera parecida a los liposomas sencillos, aunque requieren técnicas más complicadas que incluyen el uso de detergentes.

Los proteoliposomas son, estructural y funcionalmente, análogos a las membranas celulares, de las que sólo difieren por su origen y grado de complejidad. La técnica que consiste en aislar una proteína de una membrana natural e incorporarla, tras su purificación, a una bicapa lipídica de composición definida se denomina reconstitución. Se trata de un método fundamental en la biología molecular de membranas.

Descritos los principales sistemas utilizados en el estudio de la fluidez, hemos de ocuparnos de la estructura y conformación de las cadenas hidrocarbonadas de los ácidos grasos, íntimamente relacionados con la noción de fluidez. La mayoría de los lípidos de las biomembranas contienen cadenas hidrocarbonadas largas; predominan los casos de 16 a 22 carbonos. Estas cadenas hidrocarbonadas constituyen la matriz de la bicapa; la fluidez de la misma depende directamente de las características de estas cadenas: su longitud, ramificación, insaturación y otras.

La más estable para los dos carbonos contiguos de una cadena hidrocarbonada es la llamada conformación alternada *anti*, es decir, aquella en la que los sustituyentes alquílicos de estos carbonos se hallan en las posiciones más alejadas posibles. (Un sustituyente alquílico es el distinto del hidrógeno.) Para la totalidad de la cadena, la conformación más estable es aquella en la que cada carbono se encuentra en conformación alternada *anti* con respecto a sus contiguos: se dice entonces que la cadena se halla en conformación *todo-anti*. Aunque se trate de la conformación más estable, las cadenas hidrocarbonadas que forman parte de las bio-

membranas suelen adoptar, por motivos que se discutirán más adelante, conformaciones de suyo menos estables con parejas de carbonos contiguos en conformación eclipsada o en la así llamada alternada *izquierda* o *gauche* (que, de aquí en adelante, significará una conformación donde hay una rotación de 120 grados del enlace carbono-carbono con respecto a la transformación *trans*; tal rotación puede ser hacia la derecha o hacia la izquierda).

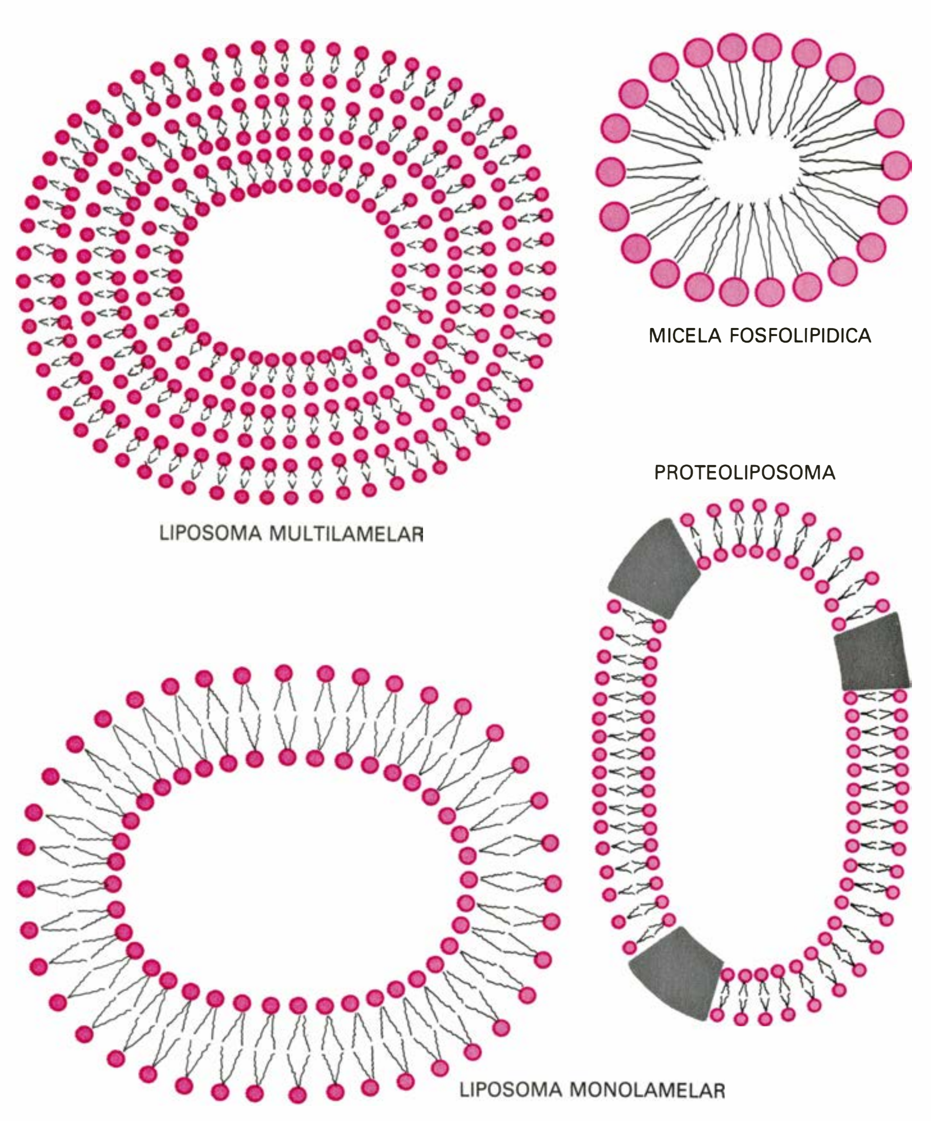
Entre las circunstancias que hacen que la conformación *todo-anti* no sea la más estable en muchos casos, citaremos las intrínsecas a la propia cadena hidrocarbonada: existencia de dobles enlaces o de sustituyentes de cualquier tipo en la misma. Es muy frecuente en las membranas biológicas encontrar lípidos anfipáticos cuyas cadenas hidrocarbonadas contienen uno, dos y hasta seis dobles enlaces. Naturalmente la rigidez que confiere el orbital enlazante π al doble enlace carbono-carbono impide cualquier cambio conformacional en la pareja de carbonos implicados; salvo contadas excepciones, los dobles enlaces de los lípidos en las biomembranas se disponen en configuración *cis*, que produce la máxima distorsión de la cadena con respecto a la conformación *todo-anti*. De igual manera, los sustituyentes que a veces se encuentran en las cadenas (metilos, hidroxilos, menos frecuentemente cicloalquilos y otros) alteran el equilibrio conformacional.

La conformación de las cadenas hidrocarbonadas adquiere una significación especial cuando se considera la ordenación o empaquetamiento de los lípidos anfipáticos en la bicapa. Una bicapa lipídica integrada por cadenas hidrocarbonadas en conformación *todo-anti* posee el grado máximo de ordenamiento; sus cabezas polares se disponen hexagonalmente. Cualquier causa que determine la aparición de confórmeros *izquierdos* o *gauche* u otros aumentará el desorden de la matriz lipídica.

La temperatura, un factor extrínseco a la cadena hidrocarbonada, condiciona el equilibrio conformacional de las cadenas hidrocarbonadas. Su aumento favorece múltiples estados conformacionales distintos del alternado *anti*. Cuando se suministra calor a una bicapa con las cadenas hidrocarbonadas en conformación *todo-anti*, el cambio que se registra en la morfología es altamente cooperativo: se da una transición de fase en la que todas las cadenas hidrocarbonadas, al alcanzar una temperatura determinada, absorben calor para pasar a un estado menos ordenado. Las fases ordenada y menos ordenada separadas por esta transición se denominan, respectivamente, gel y cristal líquido. La transición es reversible. Si rebajamos la temperatura, la fase de cristal líquido se vuelve a ordenar para producir la fase del gel con liberación de calor. Por encima de la transición de fase, las cadenas hidrocarbonadas pierden su simetría hexagonal y se disponen al azar. Además, en una bicapa desordenada los fosfolípidos rotan en torno a un eje perpendicular al plano de la bicapa y difunden lateralmente por moción térmica.

La difusión lateral de los fosfolípidos en el plano de la membrana puede medirse por espectroscopía de resonancia de espín electrónico. Esta técnica requiere modificar químicamente la molécula de fosfolípido, introduciendo un radical con un electrón no apareado. El movimiento de este radical en un medio esencialmente magnético, como es la membrana, puede detectarse a partir del ensanchamiento de las líneas espectrales que se produce por interacción espín-espín cuando dos radicales colisionan.

El coeficiente de difusión lateral de los fosfolípidos en una bicapa por encima de la temperatura de transición es del orden de 10^{-11} metros cuadrados por segundo. La facilitación de estos movimientos rotacionales y translacionales por encima de la temperatura de

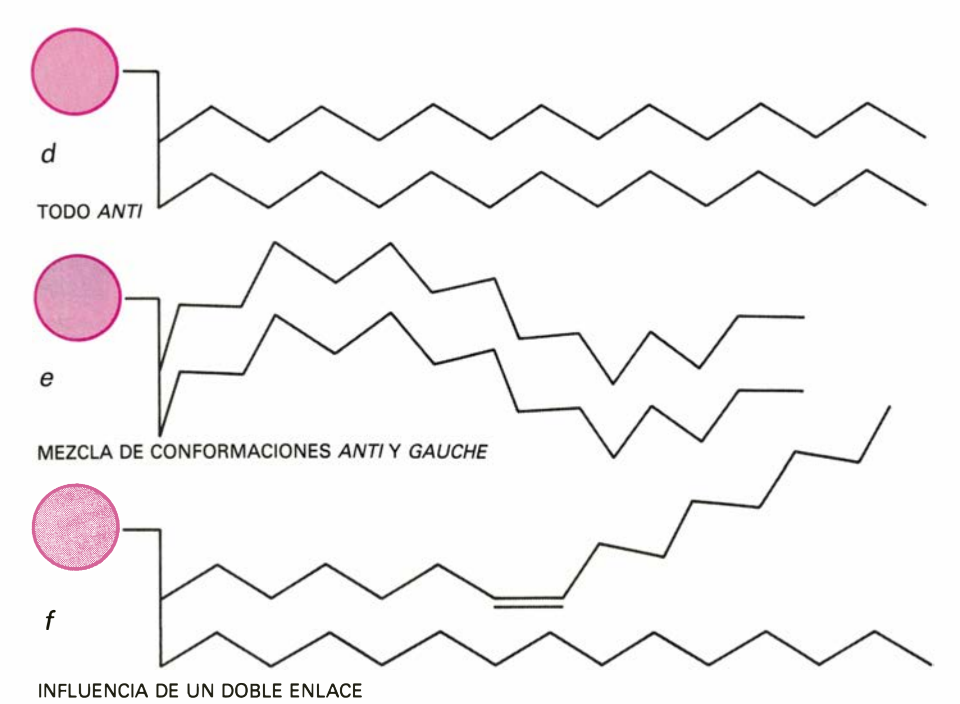


1. DISPOSICIONES ADOPTADAS POR LOS FOSFOLIPIDOS al dispersarse en agua. Los fosfolípidos son los lípidos anfipáticos más abundantes en las membranas celulares. En algunos casos su conformación puede ser la micelar. Sin embargo, los lípidos extraídos de la mayoría de las membranas celulares se dispondrán en lamelas (películas formadas por una doble capa) al dispersarse en agua. Lo mismo sucede con muchos fosfolípidos puros, como por ejemplo la fosfatidilcolina. Estas lamelas se disponen en capas concéntricas en forma de hojas de cebolla, dando origen a las estructuras llamadas liposomas. Cuando estos liposomas multilamelares (*arriba, izquierda*) se someten a irradiación ultrasónica se forman estructuras de una sola capa llamadas "vesículas lipídicas monolamelares" o liposomas monolamelares (*abajo, a la izquierda*). Existen técnicas que permiten incorporar en los liposomas proteínas intrínsecas y extrínsecas (más o menos fuertemente ligadas a la membrana) dando origen a los proteoliposomas.

transición hace que disminuya la microviscosidad, o magnitud inversa a la fluidez, de la membrana.

Debido a su composición lipídica (cadenas hidrocarbonadas de longitud variada, con dobles enlaces, ramificaciones, etcétera), las membranas celulares se encuentran a temperaturas fisiológicas en estado de cristal líquido y no suelen presentar transiciones de fase muy cooperativas. Los liposomas, formados por un solo tipo de lípido con un solo tipo de cadena hidrocarbonada, desarrollan una transición muy cooperativa con una variación de entalpía manifiesta. Los estados de gel y cristal líquido representan, respectivamente, estados de fluidez reducida y elevada. Por ser fácil el paso de gel a cristal líquido, y viceversa, estos sistemas allanan el estudio de la influencia de la fluidez en otras propiedades de la membrana. No está de más señalar aquí la diferencia entre la transición de fase de gel a cristal líquido que acontece en un sistema totalmente hidratado, como la bicapa, y la transición de fase cristalina a fase líquida del fosfolípido anhidro, tal como se observa en el punto de fusión capilar. Por ser la transición de fase de gel a cristal líquido muy cooperativa en el caso de los lípidos puros, se puede observar en la temperatura crítica, T_c, una variación de entalpía, que se detecta sin mayores dificultades por métodos calorimétricos del estilo de la calorimetría diferencial de barrido, DSC (del inglés "differential scanning calorimetry"). Las primeras observaciones de transiciones de fase en sistemas fosfolípido-agua se obtuvieron con esta técnica calorimétrica; más tarde se sacaría partido a técnicas espectroscópicas.

La DSC describe, gráficamente, la capacidad calórica a presión constante, C_p, del lípido hidratado, que, integrada con respecto a la temperatura, proporciona una medida de la variación de entalpía de la transición.



2. PARA DOS CARBONOS CONTIGUOS de una cadena hidrocarbonada de las que forman la matriz lipídica, la conformación más estable es la alternada *anti*, en la que los sustituyentes alquílicos se encuentran en las posiciones más alejadas posibles (*a*); esta conformación se ilustra con la representación en caballete (*arriba*) y la proyección de Newman (*abajo*). El aumento de temperatura y otros pueden producir conformaciones alternadas *izquierda gauche* (*b*) o eclipsadas (*c*) que de suyo son menos estables. Para la totalidad de la cadena, la conformación más estable es la *todo-anti*, en la que cada carbono está en conformación alternada *anti* con respecto a sus contiguos. Esto se ilustra en (*d*), donde se representa un fosfolípido con sus dos cadenas parafínicas en esta conformación. En (*e*) las cadenas adoptan otras conformaciones. En (*f*), finalmente, se recogen las consecuencias de un doble enlace en una cadena parafínica.

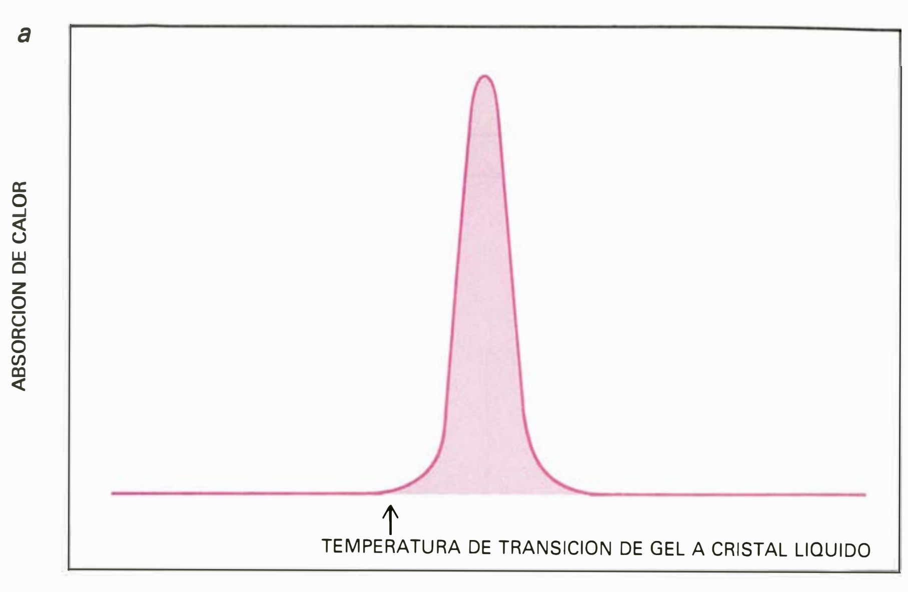
Mientras que la técnica calorimétrica nos revela sólo la existencia de una transición de fase, la técnica de difracción de rayos X contribuye a desvelar la estructura de las distintas fases en los sistemas lípido-agua y lípido-proteína-agua. La difracción de rayos X permite medir distancias repetitivas: espacios interatómicos y otros parámetros que se van reproduciendo de una forma regular en las mallas cristalinas. En las membranas biológicas, existen dos tipos principales de distancias repetitivas: la que separa las cabezas polares de las dos hojas de la bicapa (de unos 100 angström) y la que separa unas cadenas hidrocarbonadas de otras (de cuatro a cinco angström). Esta última distancia es del orden de magnitud de los parámetros habituales en cristalografía, mientras que la primera es mucho mayor. Según la ley de Bragg, el ángulo de difracción ha de ser mayor cuanto menor sea la distancia repetitiva ($n \lambda = 2d$ sen θ, donde n simboliza el orden de difracción; λ, la longitud de onda de la radiación de rayos X; d, la distancia repetitiva y θ, el ángulo de difracción). Así, la distancia entre las cadenas hidrocarbonadas puede medirse

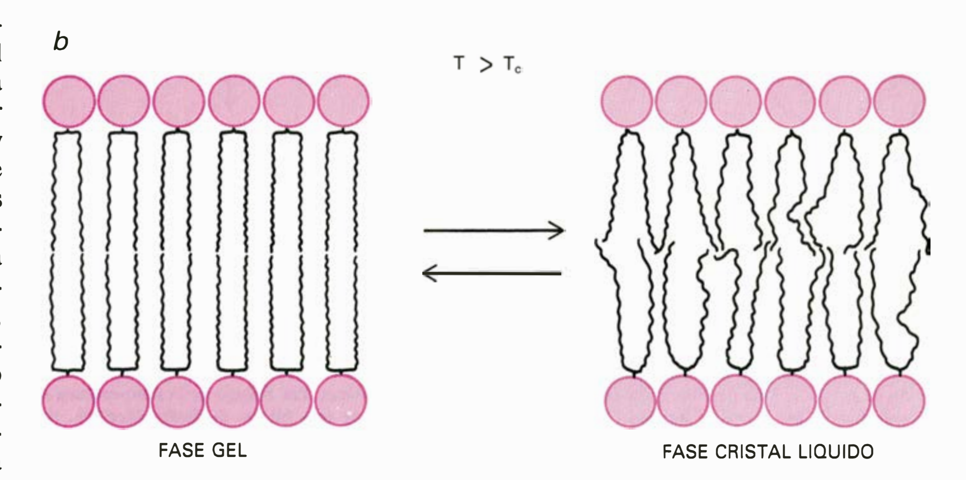
con técnicas cristalográficas de gran ángulo (que son las habituales en cristalografía), en tanto que la medición de la anchura de la bicapa requiere la utilización de técnicas de rayos X de pequeño ángulo. Con las segundas podemos medir con precisión y de forma inequívoca las dimensiones de las fases existentes al mezclar los componentes de las membranas biológicas en distintas proporciones y a distintas temperaturas.

Los resultados de estas observaciones se resumen en los diagramas de fases. Según las condiciones de la preparación, predominan la fase lamelar o de bicapa, ya comentada (L_α), otros tipos de fases lamelares (L_β, L_γ) o, en el caso de lípidos incompletamente hidratados o de algunos determinados fosfolípidos (incluso completamente hidratados), por ejemplo fosfatidiletanolamina, la llamada fase hexagonal H. Dentro de la fase lamelar (L_α), la difracción de rayos X permite distinguir entre las dos fases mencionadas de gel y cristal líquido. En los diagramas de fase se suele incluir la línea de temperaturas que separa a ambas fases (curva de coexistencia). A temperaturas superiores a las indicadas por la curva, todos los lípidos se hallan en estado desordenado, siendo la distancia media ente las cadenas hidrocarbonadas de cuatro a cinco angström. Por debajo de dicha temperatura, los lípidos empiezan a ordenarse, en cuyo caso la distancia repetitiva correspondiente es de 4,2 angström. Podemos observar, pues, la transición de fase de un lípido cuantificando la intensidad de la difracción correspondiente a 4,2 angström. Como es natural, si se trata de un lípido puro, la transición será cooperativa, y todo el lípido fundirá a la misma temperatura.

En el caso de una mezcla de fosfolípidos con distinta T_c que formen bicapas de composición compleja, aparecerá una temperatura de transición intermedia entre las de los fosfolípidos puros. Ocurre así con el dimiristil fosfatidilglicerol y dipalmitil fosfatidilglicerol, que tienen, en estado puro, una T_c de 22 y 40 grados Celsius, respectivamente; cuando se hallan mezclados, una transición difusa que comienza a 28 grados Celsius.

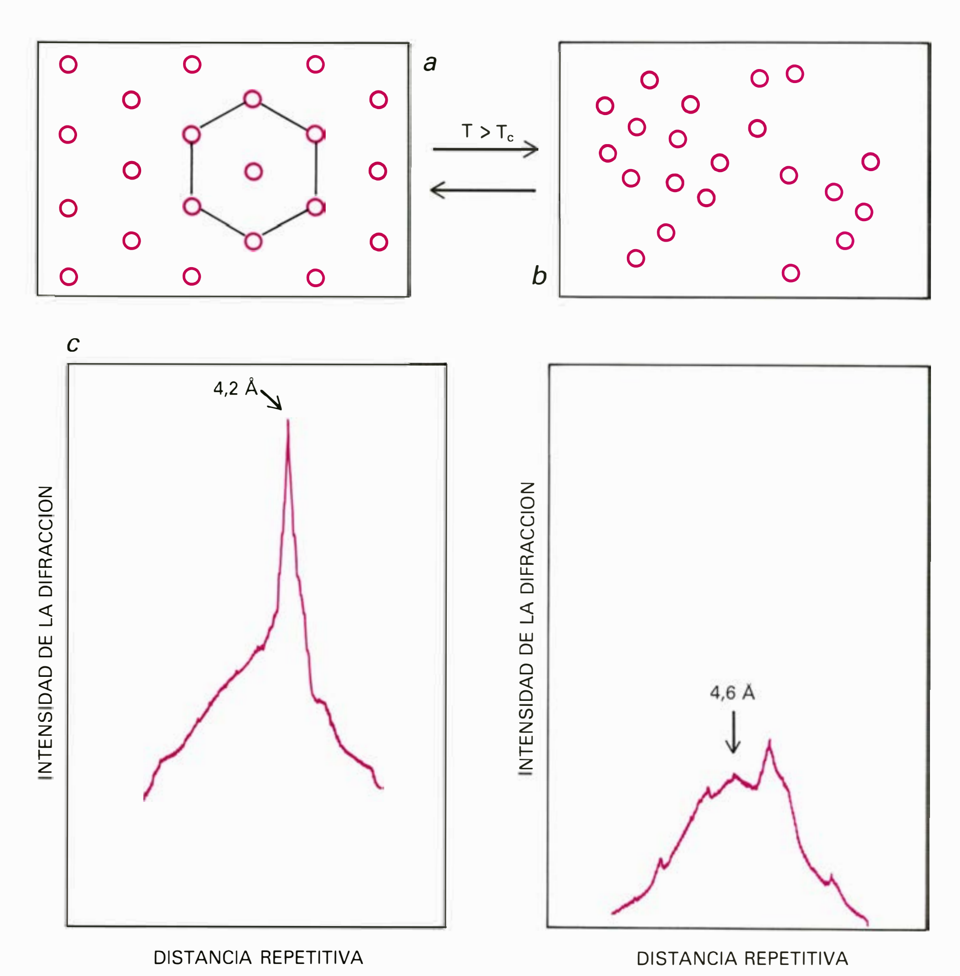
En una mezcla de lípidos se puede producir la separación de fases en el seno de la bicapa, con la aparición de zonas locales de menor fluidez por factores diversos. Por su importancia biológica, destacaremos la acción de los iones Ca^{++} en mezclas de lípidos que contienen algunos fosfolípidos cargados negativamente. Los iones Ca^{++} provocan que estos fosfolípidos ácidos se segreguen de los demás a través de una transición de fase isoterma, dando lugar a regiones de la membrana ricas en fosfolípidos ácidos, en fase gel, mientras que el resto de la membrana queda enriquecida en fosfolípidos neutros o cargados positivamente. Este efecto del Ca^{++} sobre los lípidos de las membranas biológicas se ha propuesto como un mecanismo fundamental para la fusión de membranas biológicas en procesos tales como la neurosecreción, la endocitosis y otros.





3. CRISTAL LIQUIDO Y GEL. Las bicapas lipídicas pueden presentarse en estos dos tipos de fases de fluidez muy diferente. En la fase gel, los fosfolípidos están empaquetados de manera que las cadenas hidrocarbonadas se hallan en conformación *todo anti* y las cabezas adoptan una simetría cuasi-hexagonal. En estas condiciones, la movilidad rotacional y translacional de los fosfolípidos es muy limitada. Si se suministra calor al sistema, la bicapa en fase gel sufre una transición a la fase cristal líquido. Esta transición, reversible, va acompañada de un cambio calórico detectable por calorimetría diferencial de barrido (*a*). La variación de entalpía así observada permite detectar el cambio conformacional (*b*).

Hasta ahora hemos identificado el término fluidez con la propiedad de las bicapas lipídicas por encima de la temperatura de transición (T_c), cuyas cadenas hidrocarbonadas se hallan predominantemente en conformación *izquierda* o *gauche*. Hemos distinguido entre dos estados, más o menos fluidos, es decir, líquido cristalino y gel. Sin embargo, cuando tratamos con mezclas complejas de lípidos, verbigracia membranas celulares en las que no hay transiciones cooperativas sino graduales, resulta conveniente aplicar una técnica que nos proporcione una información fácilmente relacionable con la fluidez a cada temperatura. La espectroscopía de polarización de fluorescencia cumple este requisito. Esta técnica exige la introducción, en la bicapa, de una molécula hidrofóbica fluorescente que actúa de sonda. Al incidir sobre la sonda una radiación de excitación, polarizada en un plano, el mantenimiento de esta



4. DISTANCIAS REPETITIVAS en las membranas biológicas, medidas merced a la técnica de la difracción de rayos X. Una de ellas es la distancia entre cadenas parafínicas. Por debajo de la temperatura de transición de gel a cristal líquido (T_c), las cadenas parafínicas de una bicapa de fosfatidilcolina, por ejemplo, se encuentran empaquetadas hexagonalmente. Esto se ilustra en (*a*), donde se representa una sección de la bicapa según un plano perpendicular a las cadenas parafínicas por su zona media. Los círculos simbolizan estas cadenas, que por encima de T_c, aparecen desordenadas (*b*). Representando las intensidades de las difracciones de los rayos X con relación a las distancias repetitivas, por medio de un microdensitograma, se obtiene un pico para la bicapa estudiada por debajo de T_c a 4,2 angtröm (*c*). Sin embargo, por encima de T_c, solamente se observa un máximo poco pronunciado a 4,6 angström.

polarización en la radiación emitida significará que la sonda tiene escasa libertad de movimientos. Por el contrario, la despolarización de la radiación fluorescente emitida indicará que la sonda, en el tiempo transcurrido entre la excitación y la emisión, ha variado su orientación. El grado de despolarización de la señal refleja de este modo la microviscosidad de la membrana.

La espectroscopía de polarización de fluorescencia nos permite observar la diferencia existente entre la transición de fase cooperativa de un fosfolípido puro, con un cambio brusco de microviscosidad, y el cambio gradual que se registra por efecto de la temperatura, en una mezcla de lípidos de origen natural.

Podemos ahora considerar el efecto que, sobre la fluidez de la membrana, ejercen otros componentes de la misma, que no poseen cadenas hidrocarbonadas largas. Estos componentes son fundamentalmente los esteroles, de los que el colesterol es el más abundante en mamíferos, y las proteínas. Se ha estudiado con gran detalle la acción del colesterol sobre la fluidez lipídica, tanto por debajo como por encima de la temperatura de transición de gel a cristal líquido: el colesterol perturba y modula la fluidez lipídica. Se ha visto, en estudios de calorimetría diferencial de barrido (DSC), cómo con cantidades crecientes de colesterol se va eliminando progresivamente la transición de fase. La espectroscopía de polarización de fluorescencia revela un aumento gradual de la microviscosidad al añadir cantidades crecientes de colesterol a un sistema formado por un fosfolípido puro y agua por encima de la temperatura de transición del fosfolípido puro (T_c). Lo contrario ocurre por debajo de T_c. Por encima de la temperatura de transición T_c, se da una modulación de la fluidez lipídica, haciéndose el lípido menos fluido, mientras que por debajo de T_c se impide que el lípido cristalice y aumenta, por tanto, su fluidez. Es decir, el colesterol actúa de amortiguador con relación a la fluidez de la bicapa lipídica, induciendo una fluidez intermedia.

¿Qué decir de la incidencia de las proteínas? Las proteínas extrínsecas son hidrosolubles. Algunas, como el citocromo *c*, se incorporan fácilmente a los liposomas. Basta añadir la proteína en una disolución acuosa de la fuerza iónica adecuada al lípido anhidro. Al dispersarse la mezcla, se consigue la formación inmediata de proteoliposomas.

Entre las proteínas intrínsecas mejor estudiadas se encuentran la ATPasa transportadora de Ca^{++} del retículo sarcoplásmico, la bacteriorrodopsina de *H. halobium*, la citocromo *c*-oxidasa mitocondrial, la glicoforina de la membrana eritrocitaria, la apolipoproteína de la mielina, la rodopsina de la retina ocular y algunos polipéptidos, como la gramicidina y la melitina. La incorporación de proteínas intrínsecas en los liposomas requiere la cosolubilización de proteínas y lípidos, bien en disolventes orgánicos o bien en medio acuoso en presencia de detergentes, según la técnica de reconstitución antes mencionada.

El efecto de las proteínas sobre las transiciones de fase de las bicapas lipídicas difiere según el tipo de proteínas de que se trate. Dentro de las proteínas extrínsecas o periféricas, algunas, como la ribonucleasa o el polipéptido polilisina, aumentan la variación de entalpía asociada a la transición de fase, sin modificar la T_c. Se cree que este tipo de proteínas se unen a la superficie de la bicapa sin penetrar en la región de las cadenas hidrocarbonadas. Por el contrario, proteínas extrínsecas como el citocromo *c* y la proteína básica de la mielina inducen una reducción drástica de la temperatura de transición T_c y de la entalpía asociada a la misma, $\triangle H$. Estas proteínas se unirían a la superficie de la bicapa produciendo cierta deformación o penetrando parcialmente en la misma.

Las proteínas intrínsecas estudiadas actúan de forma similar al colesterol en cuanto a la disminución y supresión de las transiciones de fase y no variación de T_c. La espectroscopía de polarización de fluorescencia indica también que el efecto de las proteínas intrínse-

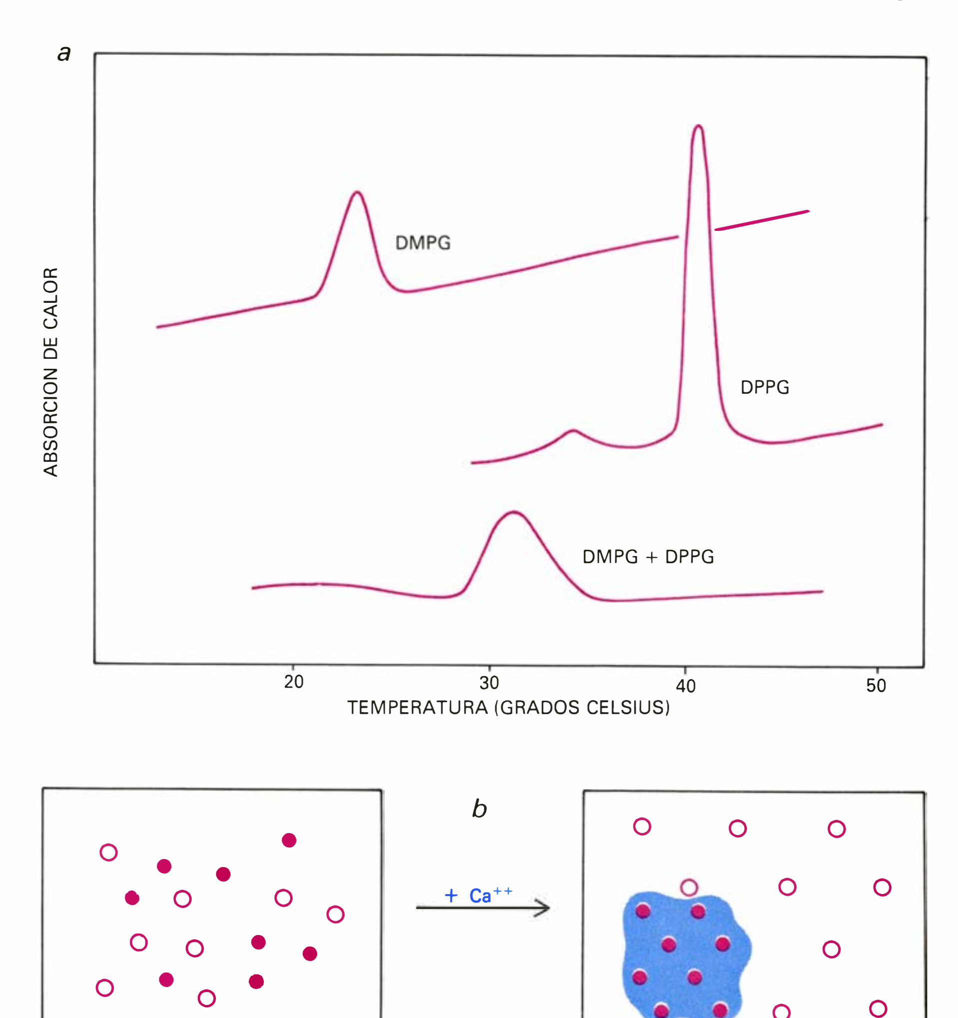
cas sobre la microviscosidad de la bicapa es similar al del colesterol, con la diferencia de que, según esta técnica, ni siquiera con las preparaciones de más alta proporción proteína-lípido que se pueden obtener sin desnaturalizar la proteína se consigue eliminar por completo la transición de fase del lípido. El estudio de la microviscosidad a unas temperaturas dadas, por encima y por debajo de la temperatura T_c del lípido puro, revela que con el lípido en fase cristal líquido la adición de cantidades crecientes de proteína produce un descenso de la microviscosidad.

En las biomembranas del tipo a que antes nos referíamos, y en las que tienen lugar transiciones de fase cuando contienen proteínas, por debajo de la temperatura de transición T_c se dan fenómenos de segregación lateral de fase con producción de zonas ricas en proteínas y otras desprovistas de ellas. Estas separaciones de fase se observan mediante la técnica de microscopía electrónica de criofractura y réplica. En esta técnica, las membranas se congelan a gran velocidad; por medio de una cuchilla se corta el bloque de hielo de la muestra, de manera que algunas de las membranas resultan fracturadas, precisamente siguiendo el plano que separa las dos hojas de la bicapa. A continuación se prepara una réplica en carbón de la cara expuesta de la membrana. Esta réplica se endurece y sombrea con platino y queda dispuesta para la observación. En estas condiciones se aprecian superficies lisas que se cree corresponden a lípidos y granulaciones de moléculas, o agregaciones de moléculas, proteicas.

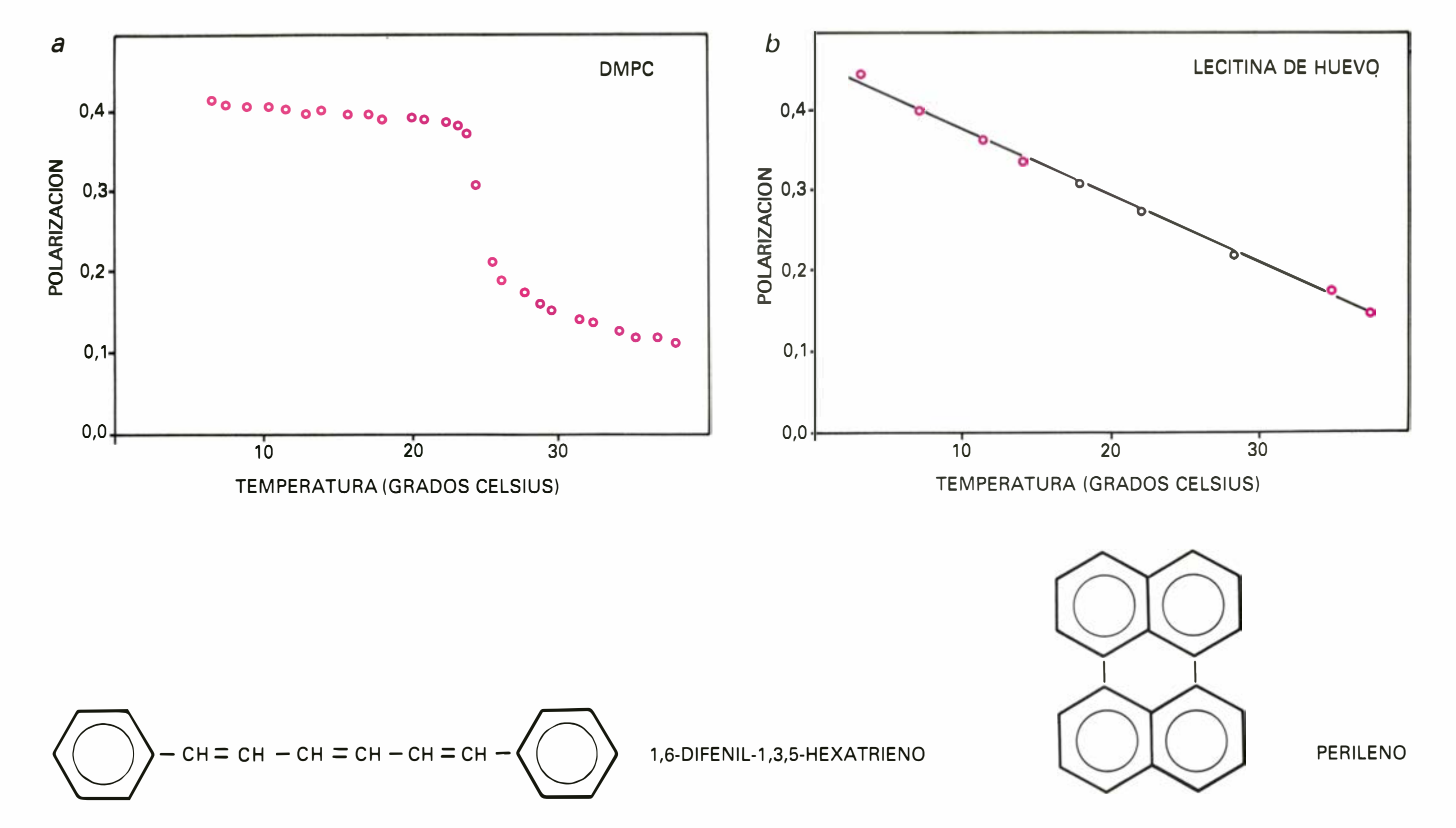
Si aplicamos la microscopía electrónica de criofractura y réplica al estudio de proteoliposomas por encima de la temperatura de transición (T_c), podemos contemplar la distribución aleatoria, en el plano de la membrana, de las granulaciones proteicas. Por debajo de T_c las proteínas se agregan en determinadas zonas de la bicapa, dejando otras áreas libres de ellas. La razón por la que se produce esta separación de proteínas estriba en que la proteína intrínseca, al intercalarse en la bicapa, influye en el proceso de cristalización. Probablemente, por su forma irregular y su tamaño no crítico (es decir que no entra en el espacio normalmente ocupado por un número entero de cadenas lipídicas), la proteína induce tensiones laterales en la estructura de la fase gel, que se traducen por un desorden de las moléculas lipídicas adyacentes. Este desorden se transmite a zonas alejadas de la proteína en forma de defectos de empaquetamiento. Si la proporción proteína-líquido es suficientemente alta, al descender la temperatura y cristalizar el lípido las proteínas irán quedando excluidas hasta constituir regiones formadas por proteínas y lípido no cristalino, con una relación proteína-lípido muy grande. Al mismo tiempo aparecerán zonas de nucleación de lípido cristalino que irán adoptando la simetría hexagonal.

En algunos sistemas membranosos que pueden existir en presencia y ausencia de colesterol sólo se produce agregación de proteínas a baja temperatura en ausencia del mismo. Tal es el caso de *Mycoplasma mycoides* var. *capri* que puede cultivarse de manera que sus membranas contengan o no colesterol. Como ya hemos dicho, éste mantiene relativamente fluida la bicapa lipídica, incluso por debajo de T_c, con lo cual el lípido no cristaliza a baja temperatura y la proteína no se segrega.

Las técnicas físicas mencionadas hasta ahora no permitían observar diferencias importantes entre el efecto del colesterol y el de las proteínas intrínsecas sobre las propiedades de la bicapa lipídica. Sin embargo, la espectroscopía de resonancia magnética nuclear (RMN) nos proporciona nueva información sobre la interacción de estas moléculas con lípidos anfipáticos de la bicapa; demuestra que el colesterol y las proteínas tienen efectos diferentes. La RMN se basa en la propiedad de orientarse dentro de un campo magnético aplicado externamente que poseen los núcleos con momento dipolar permanente (H-1, H-2, P-31, C-13, etcétera). Estos núcleos pueden cambiar sus orientaciones en presencia de un campo electromagnético, concretamente de microondas; estos cambios de energía



5. CUANDO UNA BICAPA contiene más de una especie lipídica, su transición de fase se da a una temperatura, T_c, diferente de la de sus componentes puros. Así, el dimiristil fosfatidilcolina tiene una T_c de 22 grados C, el dipalmitil fosfatidilglicerol de 39 grados C y su mezcla hacia 30 grados C, más difusa ésta última (*a*). Hay ocasiones en que una especie fosfolipídica se segrega de las demás y efectúa una transición de fase. Es el caso de ciertas mezclas que contienen fosfolípidos ácidos, cargados negativamente; se segregan de los demás y sufren una transición de fase isoterma en presencia de iones calcio (*b*).



6. ESPECTROSCOPIA DE POLARIZACION de fluorescencia: técnica que permite comparar la fluidez de una bicapa a distintas temperaturas. Para ello se requiere introducir en la bicapa una molécula fluorescente adecuada, que actúa de sonda. Ejemplos de sondas fluorescentes son el perileno y el 1,6-difenil-1,3,5-hexatrieno (DPH). La polarización de la fluorescencia de estas sondas introducidas en una bicapa formada por un fosfolípido puro desciende bruscamente cuando se alcanza la temperatura de transición del lípido T_c, como es el caso de DPH en dipalmitil fosfatidilcolina a 41 grados C (*a*). Cuando la bicapa contiene una mezcla de lípidos como la lecitina de huevo, la polarización de la sonda fluorescente, en este caso perileno, disminuye gradualmente con el aumento de la temperatura (*b*). La polarización de la fluorescencia puede relacionarse con la microviscosidad de la membrana.

son los que se detectan mediante circuitos electrónicos adecuados.

En los últimos años, se han realizado grandes avances en espectrómetros de RMN, tanto por el perfeccionamiento de sus circuitos electrónicos, como por la aplicación de electroimanes superconductores (refrigerados a la temperatura del helio líquido), que permiten aplicar superiores campos magnéticos. Estos progresos permiten obtener, hoy día, espectros muy detallados y con menor cantidad de muestra [véase "Espectroscopía de RMN de células vivas", por R. G. Shulman; Investigación y Ciencia, marzo de 1983].

Cuando se utilizan núcleos de deuterio (H-2) se obtienen espectros de separación cuadrupolar. Esta separación nos revela en qué grado de ordenamiento se halla la molécula a la que pertenece el núcleo de H-2. Preparando muestras con fosfolípidos deuterados selectivamente en uno de los carbonos de uno de los ácidos grasos, con proteínas intrínsecas (tales como la ATPasa del retículo sarcoplásmico, la citocromo *c* oxidasa, la apolipoproteína del proteolípido de mielina y algunas otras) o colesterol, se observa que, por encima de la temperatura de transición (T_c), las proteínas disminuyen el parámetro de orden "separación cuadrupolar", lo que indica un efecto de desordenación. Este efecto es más llamativo cuanto más distal se encuentra del carbono deuterado la cadena hidrocarbonada. Por otra parte el colesterol produce un gran aumento de separación cuadrupolar, lo que manifiesta ordenamiento de las cadenas. Por debajo de T_c, colesterol y proteínas impiden la cristalización del lípido.

Otro parámetro importante en la espectroscopía de RMN lo constituyen los tiempos de relajación. Se denomina relajación al proceso por el cual los núcleos que han absorbido energía de la radiación de microondas, para adoptar una orientación que supone un "estado excitado", vuelven a la orientación original. En general, las condiciones que favorecen una relajación eficaz son las que disminuyen la velocidad de movimiento del núcleo en cuestión. Esto se traduce en un aumento de la anchura de la línea de resonancia asociada con ese núcleo. Se ha observado en varias muestras de proteoliposomas preparados con lípidos específicamente deuterados, en experimentos realizados a una temperatura superior a T_c, que la anchura de las bandas de RMN de H-2 es mayor que en las bicapas de lípido puro, aunque las separaciones cuadrupolares sean más pequeñas. Estos resultados sugieren que el lípido en contacto con la proteína, además de estar muy desordenado, se intercambia rápidamente con los del resto de la bicapa, en la escala de tiempo de los experimentos de RMN de H-2 (100 micro-segundos).

La espectroscopía de RMN aplicada al isótopo P-31 se muestra de particular interés en el estudio de las membranas biológicas: el P-31 tiene una abundancia natural del 100 por ciento. Además, se puede aplicar a las mismas muestras que la RMN de H-2, incluso simultáneamente. Los resultados obtenidos al medir el tiempo de relajación (T_2), que como hemos indicado es un parámetro de orden, y utilizando las preparaciones mencionadas, revelan que las cabezas polares de los fosfolípidos interaccionan directamente con las proteínas. Como consecuencia, la movilidad de los grupos fosfato se hace más lenta y más isotrópica. Por el contrario, el colesterol actúa tan sólo como una molécula espaciadora inerte, sin que haya

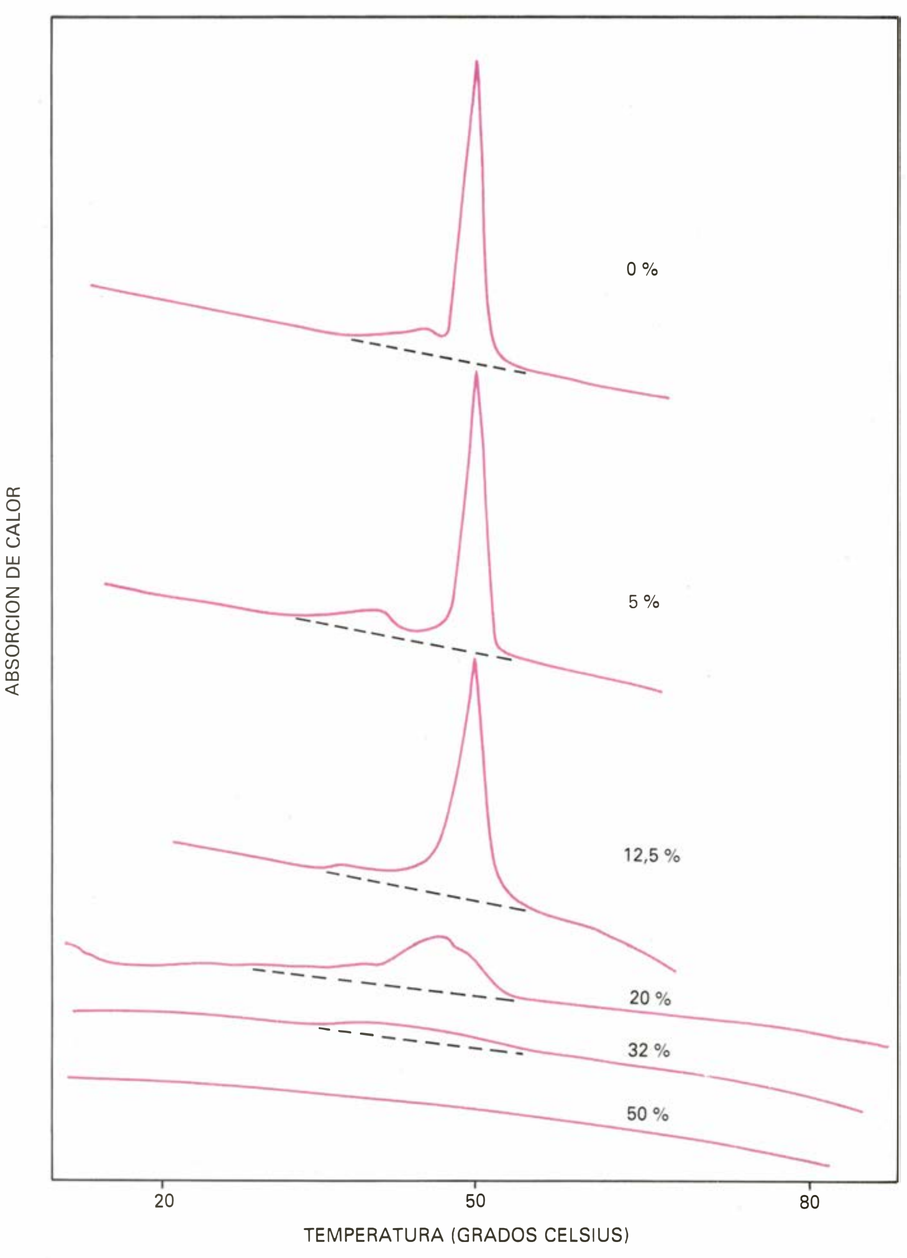
pruebas de interacción entre fosfolípidos y colesterol.

Se concluye que el principal efecto de las proteínas sobre la porción hidrocarbonada de los lípidos, tanto en modelos de membrana como en membranas nativas, consiste en desordenar las cadenas hidrocarbonadas por encima y por debajo de T_c. Esto se atribuye a la superficie irregular de la proteína inmersa en la matriz lipídica. Por el contrario, la molécula rígida y plana de colesterol aumenta el ordenamiento de las cadenas por encima de T_c, aunque las desordene por debajo de T_c.

La espectroscopía de polarización de fluorescencia nos señalaba que, a temperaturas superiores a T_c, las proteínas aumentaban la microviscosidad de la bicapa; la RMN nos puntualiza que su efecto estriba en avivar el desorden de las cadenas hidrocarbonadas. Aunque para las bicapas compuestas exclusivamente de lípidos el aumento de desorden supone una disminución de la microviscosidad, cuando se introducen proteínas intrínsecas esto no tiene por qué ser así. Recuérdese que la polarización de fluorescencia requiere la introducción de sondas en la bicapa. Las proteínas, especialmente si se hallan en concentración elevada, pueden disminuir la movilidad de la sonda fluorescente, y así dan una imagen de microviscosidad aumentada, incluso con mayor desorden de las cadenas hidrocarbonadas.

Los resultados de espectroscopía de RMN inciden en el problema de la existencia o no de un "ánulo" o zona de lípido fronterizo, inmovilizado por la proteína, por encima de la temperatura de transición del lípido puro (T_c). Estudios recientes han relacionado la existencia de este ánulo con ciertas particularidades cinéticas de enzimas intrínsecos, como la citocromo *c* oxidasa de la membrana interna mitocondrial o la ATPasa dependiente de Ca^{++} de retículo sarcoplásmico. Las pruebas de la existencia de un ánulo lipídico inmovilizado se basan en observaciones de espectroscopía de resonancia de espín electrónico. Sin embargo, los resultados de espectroscopía de RMN de H-2 no son compatibles con la existencia de dos tipos de lípido de distinta fluidez por encima de T_c, sino con la de un solo tipo.

¿Cómo reconciliar esos resultados? Teniendo en cuenta que ambas técnicas se caracterizan por distintos tiempos de correlación; en el caso de la resonancia de espín electrónico, la escala de tiempo es del orden de 10^{-8} segundos, mientras que en la RMN de H-2 se manejan tiempos de correlación de 10^{-4} segundos. Naturalmente, un lípido puede estar fijo por un espacio de tiempo de 10^{-8} segundos, pero moverse en período diez mil veces mayor. Sólo nos resta decir qué escala de tiempo, la de resonancia electrónica o la de la resonancia nuclear, nos proporciona la información adecuada en lo concerniente a sus implicaciones fisiológicas. La inmovilización del lípido por la proteína podría conducir a una modulación de la actividad enzimática de la misma; ahora bien, los ciclos catalíticos de las proteínas intrínsecas estudiadas con estas técnicas son del orden de la milésima de segundo, es decir, mucho más lentos que la posible interacción del lípido con la proteína (10^{-8} segundos). Dicho con otras palabras, en el tiempo que un enzima de la membrana utiliza para llevar a cabo un solo ciclo catalítico, los lípidos que le rodean se han intercambiado ya con los adyacentes unas cien mil veces. Así pues, no se puede mantener la idea de un ánulo lipídico rígido creando un microentorno especial a la proteína. Por encima de la tem-

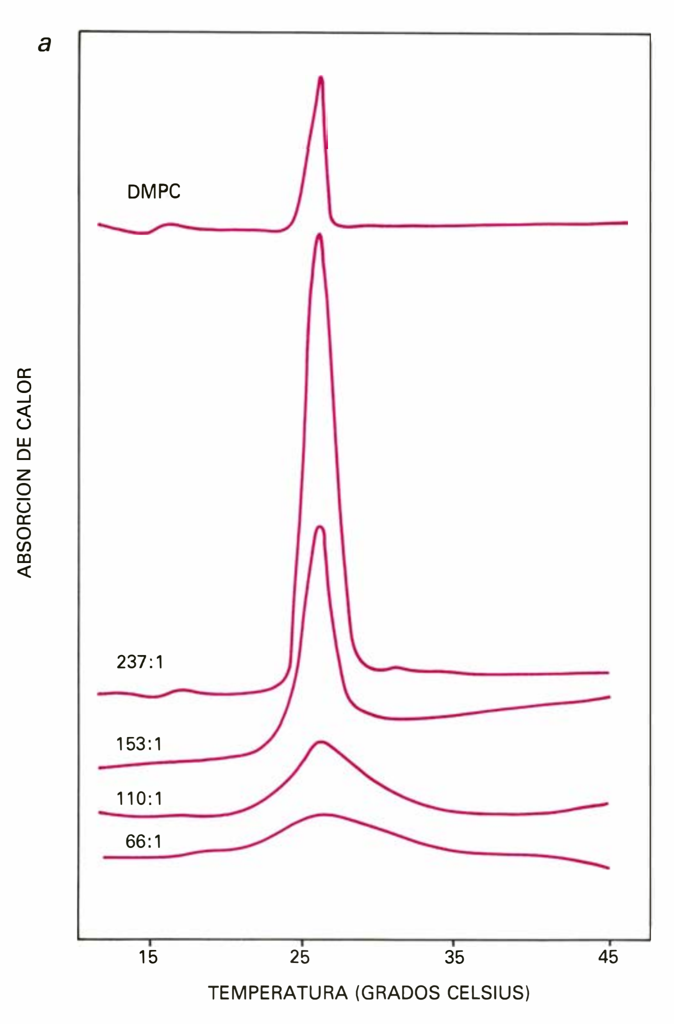


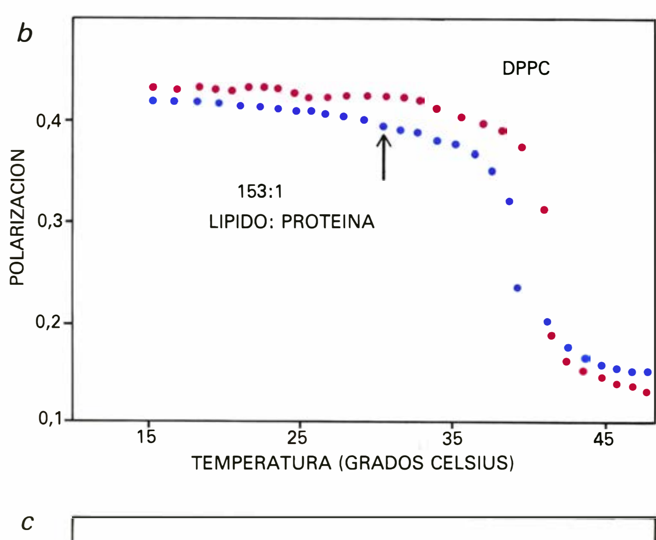
7. ESTEROLES de las membranas de las células eucarióticas. De todos ellos, el colesterol es el más abundante en mamíferos. Cuando se incluyen cantidades crecientes de colesterol en una bicapa fosfolipídica, la transición de fase se va eliminando gradualmente. Esto se ilustra en la figura donde se muestran los trazados de calorimetría diferencial de barrido de mezclas de dipalmitil fosfatidilcolina y colesterol, en distintas proporciones. Las cifras indican, en porcentaje, la proporción molar colesterol/lípidos totales. A proporción equimolar de fosfolípidos a colesterol, no se observa ya ninguna transición.

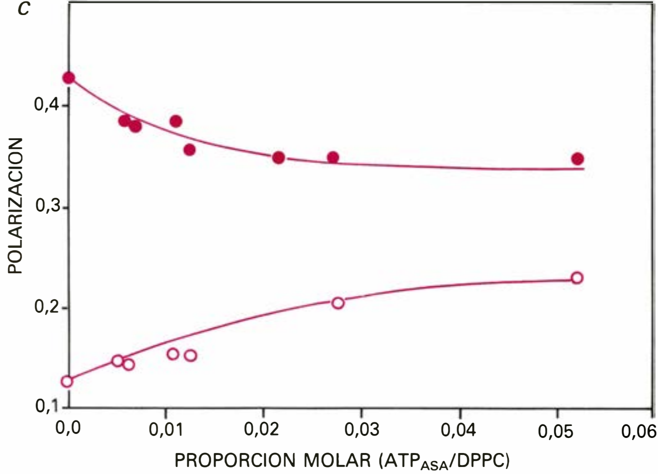
peratura de transición, la proteína no induce la aparición de una región de viscosidad distinta de la del resto de la bicapa, que a su vez module la actividad enzimática de la proteína.

Hemos visto que las características de fluidez de las membranas biológicas tienen consecuencias importantes para la segregación lateral de las proteínas. También la movilidad rotacional que poseen las proteínas intrínsecas de la membrana en el plano de ésta guarda relación con la fluidez. A su vez, de la movilidad de las proteínas dependen numerosos procesos biológicos. Entre éstos, podemos citar procesos de transporte, actividad de enzimas de membrana, transformación maligna de las células, adhesión y fusión de membranas, agregación de receptores o proteínas con función inmunitaria en respuesta a un estímulo, etcétera. De todos éstos, vamos a referirnos con mayor detalle a la actividad enzimática y su relación con la velocidad de rotación del enzima y la fluidez de la membrana. Por ejemplo, se han llevado a cabo estudios detallados sobre el tema utilizando ATPasa de retículo sarcoplásmico reconstituida con dipalmitil fosfatidilcolina. La velocidad de rotación de la ATPasa en torno a un eje, perpendicular al plano de la bicapa, se puede medir mediante fotólisis inducida por destellos de láser. El recurso a esta técnica exige que la proteína posea un cromóforo adecuado o, en su defecto, se modifique, uniéndole covalentemente uno de estos cromóforos; este último caso es el de la ATPasa de retículo sarcoplásmico, a la que se une un derivado de la eosina. La despolarización de los destellos de láser de 20 nanosegundos por el cromóforo permite valorar la velocidad de rotación del cromóforo y, por tanto, de la proteína. Examinando paralelamente los cambios en la rotación y en la actividad enzimática de la proteína con la temperatura, observamos que un aumento de esta última se acompaña de un incremento de los otros dos parámetros, como es de esperar para cualquier factor que tienda a aumentar la fluidez de la bicapa.

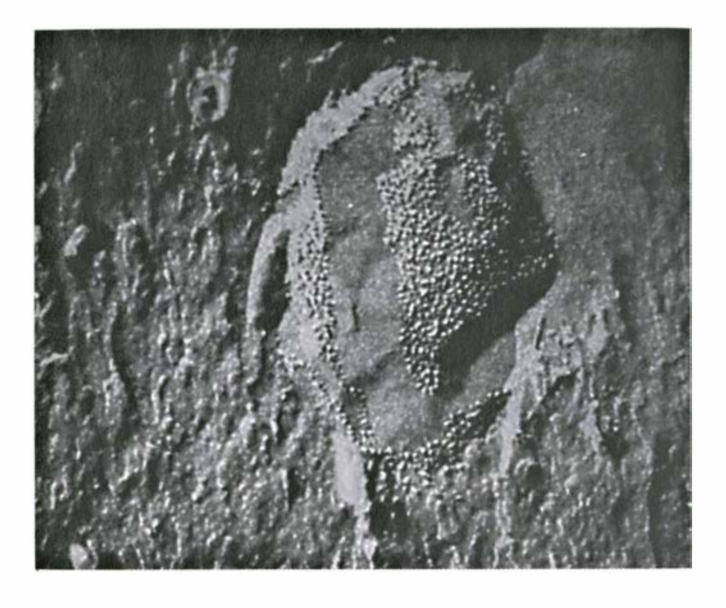
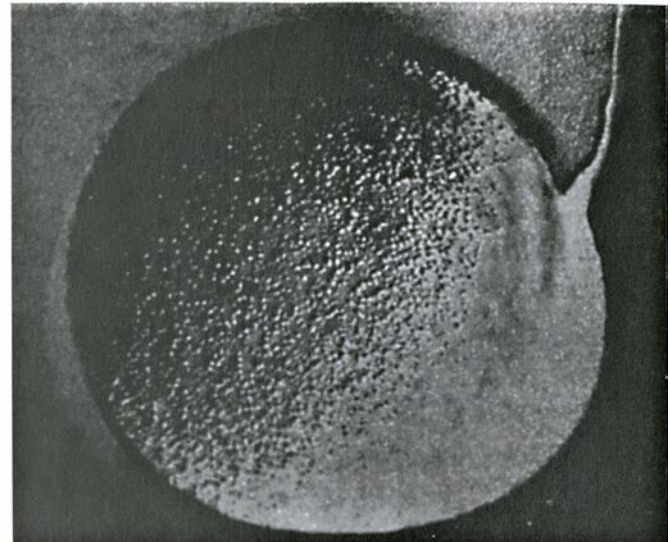
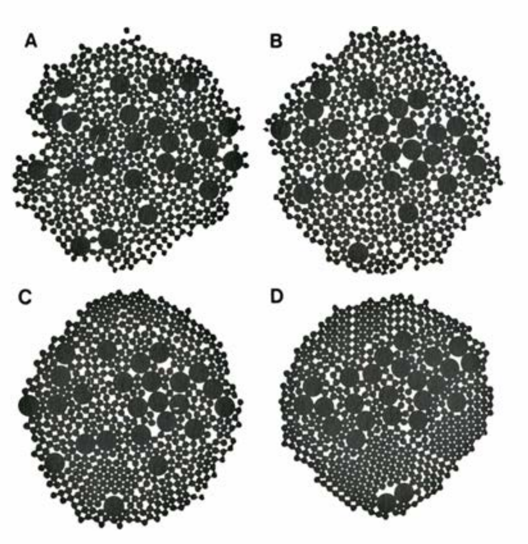
La importancia de la fluidez de las membranas, en relación con los procesos fisiológicos que acabamos de citar, explica que la evolución haya seleccionado los mecanismos adecuados para mantenerla a un nivel aceptable, a pesar de los cambios en las circunstancias ambientales: alimentación, temperatura y otros. En el caso de los microorganismos es fácil inducir su crecimiento a distintas temperaturas. Se ha demostrado que, cuando crecen a temperaturas más bajas, los lípidos de sus membra-

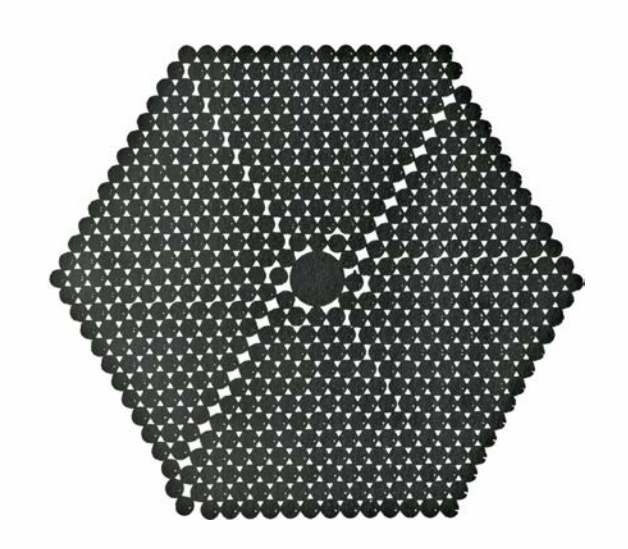






8. LAS PROTEINAS INTRINSECAS, al igual que el colesterol, disminuyen la entalpía de la transición de fase de gel a cristal líquido de los fosfolípidos con los que se asocian en las membranas. En (*a*) se muestra la reducción progresiva de la entalpía de transición de bicapas de dimiristil fosfatidilcolina al incluir cantidades crecientes de ATPasa de retículo sarcoplásmico de conejo. Las cifras en cada curva reflejan la relación molar lípido-proteína. El mismo efecto se observa en mediciones de la polarización de fluorescencia de DPH, en preparaciones de dipalmitil fosfatidilcolina y la misma proteína (*b*). Las cifras de cada curva recogen la proporción molar lípido-proteína. Las proteínas intrínsecas tienen un efecto distinto por encima y por debajo de la temperatura de transición del fosfolípido (T_c). Esto se ilustra en (*c*), donde se representan los efectos de cantidades crecientes de proteína (ATPasa de retículo sarcoplásmico) en mezclas con dipalmitil fosfatidilcolina sobre la microviscosidad medida mediante fluorescencia de polarización de DPH a 25 grados C y 46 grados, es decir, por debajo y por encima de T_c para el lípido puro. La curva superior corresponde a 25 grados C y la inferior a 46 grados C.



9. FENOMENOS DE SEGREGACION LATERAL de fases con producción de zonas ricas en proteínas y otras desprovistas de ellas por debajo de T_c, en las bicapas que contienen proteínas intrínsecas. Estas separaciones de fase se pueden observar por microscopía electrónica de criofactura y réplica. En las imágenes obtenidas por esta técnica se pueden apreciar superficies lisas que se cree que corresponden a lípidos y granulaciones de proteínas. En la fila superior podemos ver las superficies de fractura de proteoliposomas de dipalmitil fosfatidilcolina con ATPasa de retículo sarcoplásmico, rápidamente congelados desde 42 grados (*izquierda*) y 24 grados C (*derecha*) es decir por encima y por debajo de T_c (41 grados C). La razón de esta separación de fases puede estar relacionada con los defectos estructurales producidos en el empaquetamiento hexagonal de las cadenas hidrocarbonadas fosfolipídicas, en fase gel, al introducir en la bicapa una proteína de tamaño no crítico (*abajo, a la izquierda*). Abajo, a la derecha se ilustra la segregación lateral de proteínas.

nas resultan más insaturados que cuando crecen a temperaturas altas. Así, el efecto fluidificante de la temperatura elevada se compensa por el efecto contrario, de la saturación de las cadenas hidrocarbonadas. Este mantenimiento de una fluidez constante, en distintas circunstancias, se denomina adaptación homeoviscosa. No es exclusiva de los microorganismos, y se ha demostrado en todos los casos en que se ha sometido a plantas o animales poiquilotermos a procesos de aclimatación a diversas temperaturas.

Instadas por esa relación entre fluidez de las membranas y distintos procesos celulares, se han desarrollado una serie de técnicas experimentales que alterasen dicha fluidez. Las técnicas van desde la manipulación genética hasta la distorsión ambiental de las células, pasando por modificaciones nutricionales. Se puede introducir o retirar también colesterol o fosfolípidos de los sistemas membranosos. El colesterol se transfiere fácilmente de liposomas ricos a otros pobres en este lípido; se conocen proteínas capaces de transportar selectivamente determinados fosfolípidos de una membrana a otra. En el suero de mamíferos existen diversos tipos de lipoproteínas, cada una con una composición lipídica bastante definida, que pueden cumplir misiones de transporte de lípidos entre membranas. Más recientemente se han desarrollado procedimientos para modificar experimentalmente la fluidez de la membrana, hidrogenando *in situ* los dobles enlaces de los ácidos grasos de la membrana, o polimerizando por medio de los ultravioleta ácidos grasos diacetilénicos (con dos triples enlaces conjugados) incorporados a los fosfolípidos.

Es de esperar que en un futuro cercano se consiga definir magnitudes relacionadas sin ambages con la fluidez de las biomembranas. Se obtendrá entonces una correlación precisa entre las observaciones biofísicas y los datos fisiológicos. Por otra parte, el creciente grado de perfeccionamiento de las técnicas bioquímicas y biofísicas aplicadas a las biomembranas nos permitirá modular con exactitud la fluidez. No resultará difícil, en ese caso, desarrollar aplicaciones a procesos fisiológicos del mayor interés como el crecimiento, diferenciación y división celular.

Teoría reticular del confinamiento de los quarks

Se ha simulado la fuerza entre los quarks en una partícula, el protón, imponiendo una red discreta en la estructura del espacio y el tiempo. Los resultados sugieren por qué no puede aislarse un quark libre

Claudio Rebbi

El desarrollo de la mecánica cuántica ha acabado con la ciega aceptación de la idea según la cual las partículas elementales son los "bloques constituyentes" de la materia. A menudo, tales partículas no actúan en absoluto como bloques compactos, impenetrables; en muchos casos, deben entenderse como ondas. No están muy lejanos los días en que se hablaba de bloques constituyentes para designar las partículas elementales; por lo menos en el sentido de que cada partícula podía, en principio, ser aislada y observada en sí misma. Piénsese en el electrón, el protón y el neutrón: pueden separarse unos de otros y observarse como paquetes de ondas distintos. Pero incluso esta interpretación restringida o limitada del carácter de bloques constituyentes falla en el caso del quark, el supuesto constituyente del protón, el neutrón y otras partículas afines. Por lo que se nos alcanza, no hay forma de aislar los quarks. Aunque abundan las pruebas en favor de la existencia de quarks y antiquarks ligados en pares y en tripletes, nunca se ha observado un quark solo o libre.

A medida que se han ido acumulando pruebas experimentales, ha ido también tomando cuerpo la idea siguiente: si los quarks son partículas reales, deben estar permanentemente ligados dentro de las partículas nucleares. Cualquier teoría de las interacciones entre los quarks habrá de explicar este fenómeno, que se conoce por confinamiento de los mismos. Es fácil construir modelos de partículas, el protón por ejemplo, en los que los quarks constituyentes se hallan confinados. Podemos imaginar los quarks colocados en los extremos de una cuerda irrompible; son libres de moverse dentro del volumen definido por la longitud de la cuerda, aunque no pueden alejarse arbitrariamente uno de otro. Pero es trabajo titánico formular una teoría que logre explicar la ligazón permanente de los quarks y la estructura de las partículas nucleares sin violar las restricciones impuestas por la teoría de la relatividad, la mecánica cuántica y los principios de la causalidad.

Después de varios años de investigación, experimental y teórica, la mayoría de los físicos de partículas consideran que tienen por fin una teoría que da cuenta cabal de las interacciones entre los quarks. No les falta razón para esa confianza: la teoría es un análogo matemático de la teoría física más fecunda jamás desarrollada: la teoría cuántica de las interacciones en un campo electromagnético. Esta teoría se llama electrodinámica cuántica, o QED (del inglés "*Q*uantum *E*lectro*D*ynamics"), y las similitudes conceptuales de la teoría de las interacciones de los quarks con la QED se reflejan en el nombre de la nueva teoría: cromodinámica cuántica, o QCD (del inglés "*Q*uantum *C*hromo*D*ynamics").

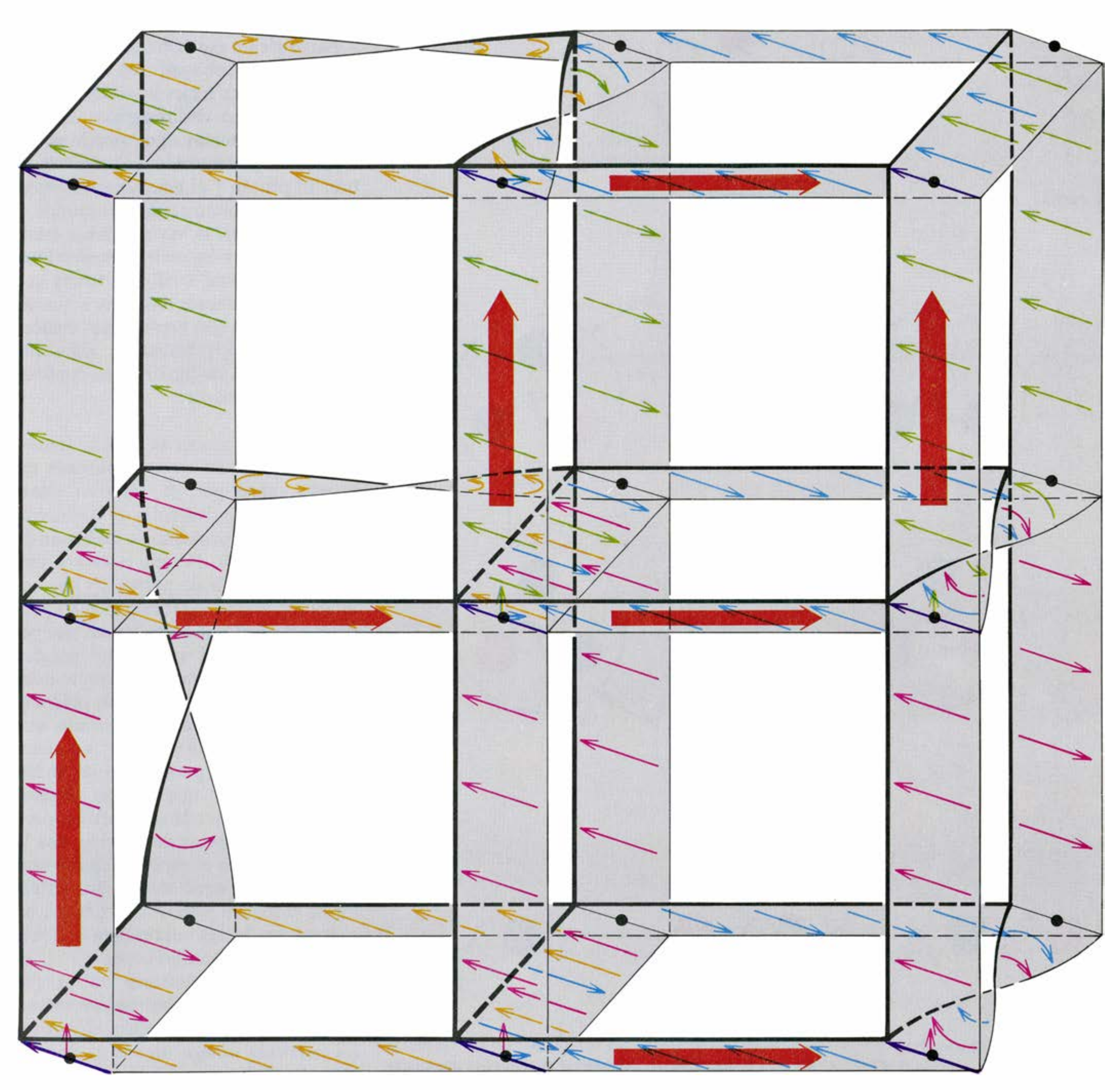
La objeción que ha retrasado la plena aceptación de la QCD reside en su complejidad matemática, que convierte cualquier predicción rigurosa y analítica de la teoría en algo excesivamente difícil. La predicción más urgentemente deseada de la QCD, la demostración del confinamiento de los quarks, no se ha logrado. Mis colegas del Brookhaven National Laboratory y yo mismo hemos aplicado métodos matemáticos, que reposan en las capacidades de un computador digital, al problema del confinamiento, y hemos logrado abrir una vía numérica hacia la solución; sin embargo, debido a que el método analiza las implicaciones de la QCD realizando series de aproximaciones de exactitud creciente, los resultados de los cálculos no presentan la misma fuerza que las deducciones lógicas basadas en primeros principios aceptados. No obstante, los resultados numéricos han aportado una prueba sólida del confinamiento de los quarks.

El marco donde encuadrar los cálculos obedece a una sugerencia que ya en 1974 realizara Kenneth G. Wilson, de la Universidad de Cornell. Wilson propuso formular la QCD en una red cúbica, una ordenación que divide el espacio y el tiempo en puntos discretos. La red es sólo una aproximación al espacio-tiempo real, pero permite realizar cálculos que, sin este recurso, serían imposibles. Cuando el entrelazado de la red se hace progresivamente más fino, los valores de las cantidades físicas definidas en la red convergen hacia los valores que la QCD prediría para ellas en el espacio-tiempo ordinario, continuo. Nuestras aproximaciones numéricas muestran que, para una red extremadamente fina, el confinamiento es una consecuencia de la QCD. Por razones que aclararemos, tanto la QCD como la QED se llaman teorías de aforo (o de contraste, en inglés "gauge theories"); el método computacional que describiré será, por tanto, una teoría de aforo en una red (en inglés "lattice gauge theory").

El modelo de quarks recibió su impulso original de la misma necesidad que había de poner orden en el gran número de partículas que presentaban interacciones fuertes, o en otras palabras, que estaban sujetas a la fuerza fuerte. El protón y el neutrón son miembros de esta clase; la fuerza fuerte los liga en el núcleo atómico. La existencia de muchas otras partículas que interaccionaban fuertemente se ha deducido a partir de los productos de desintegración de las colisiones en los aceleradores. La mayoría de tales partícu-

las viven durante un tiempo brevísimo, del orden de 10^{-24} segundos, antes de que se desintegren en otras. Todas las partículas sujetas a la fuerza fuerte se llaman hadrones, del adjetivo griego ἁδρός, que significa robusto o fuertemente constituido.

En 1962, Murray Gell-Mann, del Instituto de Tecnología de California (Cal Tech), y Yuval Ne'eman, de la Universidad de Tel-Aviv, propusieron un esquema para clasificar los hadrones en diagramas simétricos. El esquema se basaba en la teoría matemática de grupos; se llamó la octava vía. Poco tiempo después Gell-Mann e, independientemente, George Zweig, también del Cal Tech, propusieron una interpretación física de la octava vía. La clasificación matemática podía explicarse suponiendo que todos los hadrones estaban formados por otros constituyentes más fundamentales, a los que Gell-Mann puso el nombre de quarks.



1. RED DE PUNTOS: una manera esquemática de representar el espacio y el tiempo. Con una tal red los físicos pretenden comprender el campo de fuerzas que da lugar al confinamiento permanente de los quarks. Los quarks y los antiquarks, considerados los constituyentes del neutrón y del protón y de todas las demás partículas sujetas a la interacción nuclear fuerte, nunca se han observado aislados. Se agrupan en tripletes o pares. El campo responsable del confinamiento se conoce como campo cromoeléctrico; matemáticamente, se clasifica como un campo de aforo. Un aforo es análogo a una regla graduada o a un puntero; se usa para comparar cantidades físicas definidas en diferentes puntos de la red. La longitud de la regla y la dirección del puntero pueden cambiar al moverse en el espacio-tiempo. Las partículas están en los puntos, o vértices, de la red. Aquí las variables que especifican el estado de una partícula se representan por flechas violetas, que deben apuntar en una de dos direcciones en cada vértice en el primer plano de la red. Para comparar las orientaciones de dos flechas moradas, una de ellas debe llevarse al lado de la otra. Un conjunto de normas, representadas en la ilustración por las bandas que unen vértices vecinos, deben definirse para especificar los cambios hechos en la orientación de las flechas durante el transporte. El conjunto de reglas es un campo de aforo; permite que una flecha sea transportada hacia arriba o hacia la derecha de un vértice, alrededor de un grupo cuadrado de cuatro vértices llamado plaquita y volver a su punto de partida. (Las flechas transportadas son de colores, verde, amarillo, azul o rosa.) El conjunto de bandas que vuelve la flecha transportada a su orientación original representa el estado de menor energía del campo de aforo. Un conjunto de bandas que parezca una superficie de Möbius no puede enderezarse: una flecha transportada vuelve a su vértice de partida con orientación opuesta. La energía del campo de aforo definida en la red se almacena en tales plaquitas. Las gruesas flechas rojas representan las líneas de fuerza a que el campo de aforo ("gauge") da lugar.

En aquel tiempo, cada hadrón conocido podía interpretarse como combinación de tres quarks básicos (y sus correspondientes antiquarks): el quark "arriba" (*up*) o quark *u*, el quark "abajo" (*down*) o quark *d* y el quark "extraño" (*strange*) o quark *s*. El protón es combinación de dos quarks *u* y un quark *d*; el neutrón lo es de un quark *u* y dos quarks *d*, el mesón pi de carga positiva es una combinación de un quark *u* y un antiquark *d*. Desde que se propuso la hipótesis de los quarks se han descubierto nuevos hadrones y hubo que añadir, al catálogo de partículas elementales, otros dos quarks más, por lo menos: el "encantado" (*charm*) o quark *c* y el quark "fondo" (*bottom*) o quark *b*. El modelo quark de las partículas elementales permanece como un esquema de clasificación muy acertado. Se conocen más de 100 hadrones, y todos encuentran su razón de ser en ese modelo.

No obstante ese éxito, ciertas propiedades que se atribuyeron en un comienzo a los quarks hicieron difícil aceptar la realidad física de los mismos. El problema fundamental sigue siendo el del fracaso en la detección de un quark libre. El protón y el neutrón están fuertemente ligados en el núcleo atómico, a pesar de lo cual si hay suficiente energía en una colisión nuclear pueden liberarse. Por contra, cualquier teoría que describa las interacciones entre quarks no sólo tendrá que explicar sus enlaces para formar los hadrones, sino que también habrá de justificar un confinamiento permanente.

Había otra espina no mucho menos molesta: en ciertos hadrones los quarks constituyentes parecían violar un principio fundamental de la mecánica cuántica, el principio de exclusión de Wolfgang Pauli. Se aplica éste a una amplia categoría de partículas que cobija a los quarks; establece que dos partículas de ésas que se hallen en una pequeña región del espacio no pueden ocupar simultáneamente el mismo estado mecánico-cuántico. En la práctica, el principio de exclusión impide que dos quarks de la misma clase, por ejemplo dos quarks *u*, puedan formar un hadrón, a menos que tengan espines opuestos. El espín de un quark es como el momento angular de la tierra, con la salvedad de que el espín del quark está cuantificado: puede tomar sólo un valor de entre dos posibles. Por tanto, en todo grupo de tres quarks habrá, al menos, dos con el mismo espín.

Existen varios hadrones en los que tres quarks idénticos deben aproximarse entre sí lo suficiente para enlazarse. La partícula omega menos es uno de ellos. Predicha su realidad por el modelo quark, su posterior descubrimiento en 1964 por Nicholas P. Samios, Ralph P. Shutt y sus colaboradores en Brookhaven ofreció un fuerte respaldo a ese marco teórico. Ocurría, además, que la omega menos resultaba bastante intrigante ya que el modelo quark predecía que debían formarla tres quarks *s*, cuya proximidad parecía violar el principio de exclusión. Por estas y otras razones, los físicos prefirieron, en un comienzo, considerar a los quarks un útil artificio

OMEGA MENOS

LAMBDA

MESON PI POSITIVO

2. ES EL COLOR una propiedad mecánico-cuántica de los quarks que se introdujo para reconciliar el modelo de quarks con el principio de exclusión de Wolfgang Pauli. Establece el principio que dos quarks situados en una misma pequeña región del espacio no pueden ocupar el mismo estado mecánico-cuántico. Antes de la introducción de la hipótesis del color, el modelo de quarks parecía predecir la existencia de partículas en las que se violaba el principio. Así, los constituyentes de la partícula omega menos son tres quarks extraños, o quarks *s*, que se los suponía en el mismo estado mecánico-cuántico. La paradoja se resuelve admitiendo que cada uno de los tres quarks *s* es de uno de tres colores: rojo, violeta y verde. Al ser de color distinto, los quarks se hallan en diferentes estados mecánico-cuánticos y se salva así el principio de exclusión. El color nunca se observa aislado, de manera que los quarks coloreados deben combinarse de tal manera que la partícula omega menos sea incolora (o blanca). La hipótesis del color establece, además, que todas las combinaciones posibles de los colores de los quarks son igualmente probables, siempre que la partícula compuesta permanezca incolora. Por ejemplo, la partícula lambda, que está formada por tres quarks, puede ocupar cualquiera de seis estados incoloros con igual probabilidad, a saber, las seis permutaciones de los tres colores rojo, violeta y verde. El mesón pi positivo, que consta de un quark y un antiquark, puede ocupar cualquiera de tres estados incoloros con igual probabilidad: rojo y azul (antirrojo), verde y magenta (antiverde) o violeta y amarillo (antivioleta). El confinamiento de los quarks y del color se representa mostrando aquéllos unidos por una cuerda irrompible. Los quarks pueden moverse dentro de la partícula siempre que permanezcan dentro de los límites de la cuerda.

matemático; la cuestión de su existencia física se dejó, momentáneamente, de lado.

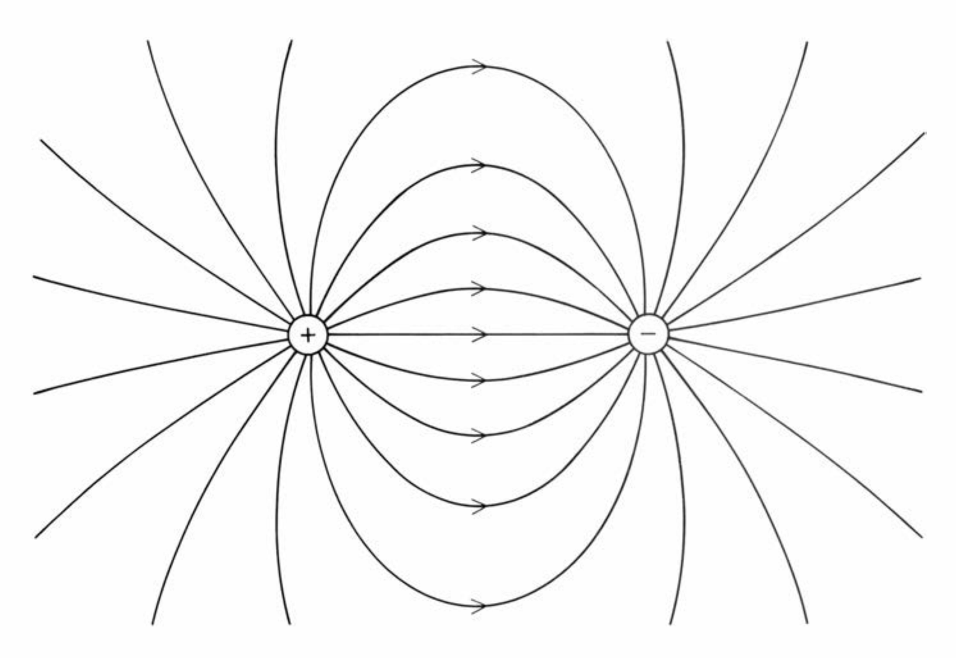
El modelo quark y el principio de exclusión se dieron la mano tras el trabajo teórico de Oscar W. Greenberg, de la Universidad de Maryland en College Park, por un lado, y el de Moo-Young Han, de la Universidad de Duke, y Yoichiro Nambu, de la Universidad de Chicago, por otro. ¿De qué modo se opera la reconciliación? Suponiendo que cada clase de quark existe en tres formas. Si un quark *s* en el estado *A* se combina con un quark *s* en el estado *B* y un quark *s* en el estado *C* para constituir la partícula omega menos, queda salvado el principio de exclusión. A la hora de designar los estados de un quark, los físicos usan el nombre de colores; un quark puede aparecer en rojo, violeta y verde; los antiquarks lo harán en los tres colores complementarios azul (antirrojo), amarillo (antivioleta) y púrpura (antiverde). El prefijo "cromo" en cromodinámica cuántica remite a la terminología de colores empleada, de una manera arbitraria por supuesto.

La introducción del color necesitaba complementarse con otra hipótesis, si quería mantenerse la acertada clasificación de los hadrones. Aunque el nuevo grado de libertad del color extendía el modelo de quarks hasta partículas del tipo de la omega menos, también provocaba una multiplicación de los hadrones. En efecto, la partícula lambda está formada por un quark *u*, un *d* y un *s*; si cada quark puede existir en cualquiera de los tres colores, habría, diríase, nueve partículas lambda, una para cada combinación de colores; pero sólo observamos una. Para evitar tal redundancia, se añade la hipótesis según la cual los quarks que integran un hadrón sólo pueden tener las combinaciones de colores que dejan al hadrón incoloro, o blanco, si se suponen las reglas de suma de colores (de la luz ordinaria). Los tres quarks de un protón, o una partícula lambda, deben incluir uno rojo, otro violeta y un tercero verde, mientras que el quark y el antiquark de un mesón pi serán rojo y azul, o violeta y amarillo o verde y púrpura. Como el color "total" es siempre el mismo, en el sentido mecánico-cuántico de que cada estado incoloro puede presentarse con la misma probabilidad, existirá una sola partícula lambda y un mesón pi positivo, nada más.

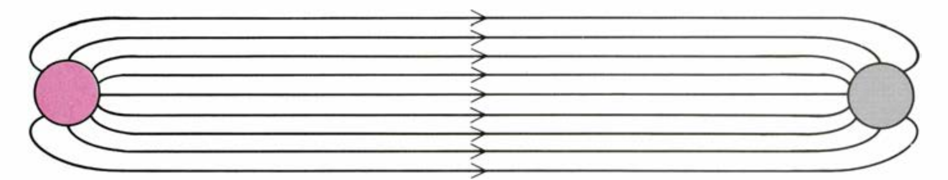
A finales del decenio de 1960, varios ensayos experimentales habían venido cosechando pruebas sólidas de que los quarks constituyentes de los hadrones eran partículas reales, y no sólo meros constructos matemáticos. De notable importancia resultó una serie de experimentos realizados en el Stanford Linear Accelerator Center (SLAC) por Jerome I. Friedman y Henry W. Kendall, del Instituto de Tecnología de Massachusetts, y Richard E. Taylor, del propio SLAC. Se lanzaron electrones de alta energía contra un blanco fijo de protones para ver si los protones tenían alguna estructura interna. Al examinar los productos de desintegración de las colisiones, quedó demostrado que en el interior de los protones había constituyentes con todas las propiedades atribuidas a los quarks.

No se detectaron los quarks libres, es cierto. Pero los experimentos mostraron que los quarks se movían dentro del protón con bastante libertad. El resultado tenía su punto de paradoja: ¿cómo podían unas fuerzas, capaces de retener los quarks permanentemente ligados entre sí, permitirles también moverse con cierta soltura cuando están en contacto inmediato dentro del protón? En este hadrón, los tres objetos con apariencia de quark quedan ligados entre sí, a imagen de las tres bolas de las boleadoras de los gauchos. Las bolas se mueven libremente siempre que permanezcan dentro de los límites de la cuerda que las une; la cuerda evita que se separen más.

Este nuevo dato experimental de existencia de los quarks, combinado con la introducción del color, supuso un gran paso hacia la formulación de



3. LINEAS DE FUERZA del campo electromagnético, que se esparcen por el espacio. El campo que se muestra es el producido por dos partículas de carga eléctrica opuesta; aunque las líneas de fuerza son más densas en la región entre las partículas, también se extienden en otras direcciones. La intensidad del campo en un punto (es decir, la intensidad de la fuerza "sentida" en un cierto punto por una carga eléctrica unidad) es proporcional al número de líneas que cruzan una superficie de área unidad, ortogonal a las líneas de fuerza, que pasa por dicho punto. La fuerza electromagnética que se genera por una única fuente de carga eléctrica disminuye en una razón del cuadrado de la distancia a esa misma fuente.



4. COMPRESION DE LAS LINEAS DE FUERZA entre dos partículas dentro de un tubo delgado de sección recta uniforme; convertiría la fuerza entre las partículas en constante, con independencia de la distancia que medie entre ellas. Una superficie de área unidad ortogonal al tubo cortaría siempre el mismo número de líneas de fuerza, sin importar dónde se colocase la superficie a lo largo del tubo. Como la fuerza que liga las partículas permanece constante, el aumento de la separación entre las partículas en una cierta cantidad exigirá siempre la misma cantidad de energía, con independencia de cuán alejadas estuvieran al empezar. Sería necesaria una cantidad infinita de energía para liberar una de las partículas de la otra. Una tal compresión de las líneas de fuerza conduciría, por tanto, al confinamiento permanente de las dos partículas. Si el radio del tubo es despreciable, el haz de líneas de fuerza se parece a una cuerda. Un modelo que explicaba el confinamiento de los quarks basado en este principio fue formulado matemáticamente en 1968 por Gabriele Veneziano, de la Organización Europea para la Investigación Nuclear (CERN), y fue interpretado como una cuerda por Yoichiro Nambu, de la Universidad de Chicago. La idea de que la cuerda pudiera comportarse físicamente como un haz de líneas de fuerza fue propuesta, en el año 1973, por Holger B. Nielsen y Paul Olesen, del Instituto Niels Bohr de Copenhague.

una teoría de la dinámica de los quarks. El color podía servir de fuente de un nuevo campo: el llamado campo cromoeléctrico, que daría lugar a una nueva clase de interacción entre las partículas cromáticas. En 1973, H. David Politzer, del Cal Tech, por un lado, y David Gross, de la Universidad de Princeton, y Frank Wilczek, de la de California en Santa Bárbara, por otro, constataron que una interacción dinámica basada en el campo cromoeléctrico conduciría a una fuerza entre los quarks progresivamente más débil a medida que se aproximaran entre sí. La predicción explicaba el movimiento casi libre de los quarks dentro del protón que se había observado en los experimentos del SLAC. Se conjeturó entonces que la misma interacción podía ser la causa del confinamiento de los quarks, aunque en aquel tiempo no había resultados teóricos que sustentaran una idea tan atractiva. Ahora bien, en cuanto se introduce el campo cromoeléctrico, se entiende que el confinamiento observado de los quarks y la hipótesis un tanto ad hoc según la cual todos los hadrones son incoloros constituyen dos aspectos del mismo fenómeno. Si un quark, por ejemplo rojo, se separa de un hadrón, el quark y el fragmento que queda se mantienen coloreados. Si el color, como la carga eléctrica, es la fuente del campo, puede haber una fuerza de atracción entre los dos fragmentos coloreados. El confinamiento de los quarks puede aparecer si la atracción entre los dos fragmentos es tan fuerte que resulta imposible separarlos más allá de cierto límite.

En los primeros años de la década de 1970 se desarrolló un modelo dinámico, independiente del modelo quark, para justificar ciertas propiedades de los hadrones que no acababan de recibir explicación satisfactoria en este último. Según el modelo dinámico, el hadrón no era un punto o una partícula esférica; la figura de la cuerda lo describiría mejor. La cuerda gira o vibra según los modos prescritos por las leyes de la dinámica relativista; se exige, además, que sus extremos se muevan a la velocidad de la luz. Los cálculos muestran que la fuerza actuante a lo largo de la cuerda debe ser enorme: unas 14 toneladas. Las vibraciones cuantificadas de esa tensa cuerda pueden identificarse con ciertos hadrones.

El modelo de la cuerda del hadrón y la analogía de las bolas pueden combinarse. Si tres quarks, o un quark y un antiquark, se colocan en los extremos de la cuerda relativista, la tensión de ésta puede explicar el enlace permanente de las partículas. Pero el modelo de la cuerda, como el concepto original de quark, es de suyo una mera abstracción matemática: la cuerda es un objeto unidimensional. ¿Podría, sin embargo, ser una aproximación de alguna otra estructura que resultara más aceptable desde un punto de vista físico?

En 1973 Holger B. Nielsen y Paul Olesen, del Instituto Niels Bohr de Copenhague, indicaron que la cuerda podía interpretarse como un haz de líneas de fuerza de un campo. En un campo electromagnético, las líneas de fuerza adquieren las formas familiares de las limaduras de hierro en una hoja de papel colocada sobre un imán. La intensidad del campo es proporcional a la densidad de las líneas de fuerza. Cuando las líneas de fuerza se separan, como sucede en los puntos cada vez más alejados de los polos de un imán ordinario, disminuye la intensidad del campo. Si las líneas de fuerza se encierran dentro de un tubo de sección uniforme, la intensidad del campo permanece constante en toda la longitud del tubo. La fuerza necesaria para separar un quark y un antiquark en los extremos opuestos de un tal tubo permanecería también constante, con independencia de cuán alejadas se hallaran ambas partículas. Para liberar uno de los quarks, habría que aportar una cantidad infinita de energía.

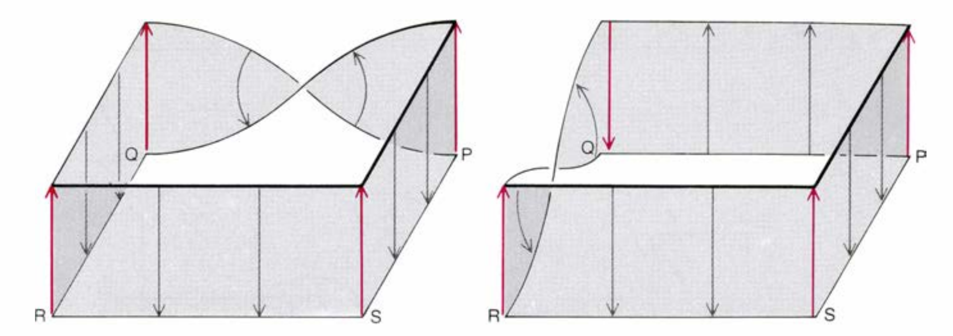
La realidad mecánico-cuántica de los quarks y las cuerdas exige que las líneas de fuerza asociadas con la interacción del color de los quarks actúen de una manera muy diferente de como lo hacen las líneas de fuerza asociadas con la interacción electromagnética de las partículas dotadas de carga eléctrica. Por propagarse una y otra fuerza en el vacío, más de uno supondría que cualquier diferencia entre ellas podría obedecer a la naturaleza intrínseca de las propias fuerzas y no a la interacción de las fuerzas con el vacío. En mecánica clásica, o newtoniana, tal hipótesis sería correcta. Pero la verdad es que no puede haber interacción entre un campo y el vacío clásico: por definición, el vacío clásico es un estado que no contiene materia ni energía. En mecánica cuántica, hasta el vacío tiene una estructura capaz de alterar la propagación de los campos y las fuerzas.

La estructura del vacío es una conse-

A B C D
A B C D
A B C D

5. SENTIDOS DE DOS FLECHAS definidos sólo en los vértices de una red: no pueden compararse, a menos que haya algún modo de transportar las flechas de un vértice al otro. En el diagrama superior, las flechas de los vértices *A* y *B* parecen apuntar en el mismo sentido, mientras que las flechas en *C* y *D* apuntarían en sentidos opuestos. Sin embargo, la comparación presupone la existencia del papel en el que el diagrama está dibujado, que actúa de intermediario sobre el que una flecha puede trasladarse al lado de la otra. Podríamos suponer que cada par de puntos estuviesen conectados por una cinta de papel semi-torcida, como en el diagrama central; entonces la comparación da el resultado contrario. Sin un campo de aforo que proporcione una norma de transporte de un punto a otro, las orientaciones de las flechas en puntos adyacentes no pueden compararse. El diagrama inferior muestra que el resultado de una comparación no cambia cuando el sentido de una flecha se invierte en cualquier punto; la comparación permanece válida si la cinta se tuerce o se endereza a la vez. La propiedad de los campos de aforo que permite compensar un cambio en el sentido de cualquier flecha se llama invariancia de aforo local.

cuencia del principio de incertidumbre de Werner Heisenberg. Establece una versión de dicho principio que, en cualquier suceso físico, no se conoce con absoluta certeza la energía liberada durante el suceso, incertidumbre que está relacionada con la imposibilidad de decidir en qué tiempo exacto ocurrió. De un modo más ajustado: el producto de la incertidumbre respecto a la energía y la incertidumbre respecto al tiempo no es menor que cierta constante numérica. Para un suceso restringido a un intervalo extremadamente corto, la incertidumbre respecto a su energía adquiere, en correspondencia, un valor grande. Por tanto, en cualquier intervalo corto habrá una probabilidad apreciable de que el vacío mecánico-cuántico tenga una energía no nula.



6. PLAQUITA FRUSTRADA es aquella que tiene un torcimiento unidad en alguno de los cuatro campos de aforo que forman sus lados. Los campos de aforo, representados por cintas, pueden torcerse o enderezarse localmente para compensar la inversión de una flecha en algún punto, pero el torcimiento total de las cuatro cintas de la plaquita no puede eliminarse. El torcimiento entre *P* y *Q* (*izquierda*) se elimina cambiando el sentido de la flecha en *Q*. El efecto producido se ciñe a transferir el torcimiento a la parte de la cinta entre *Q* y *R* (*derecha*). Independientemente de qué flechas se inviertan, o en otras palabras de cuántas transformaciones de aforo locales se hagan, debe permanecer torcido un número impar de las cuatro cintas. Por tanto, la plaquita se mantiene frustrada a pesar de la invariancia de aforo local.

Esta energía del vacío puede manifestarse en la creación o aniquilación espontáneas de una partícula y su antipartícula, o en la aparición y desaparición de un campo eléctrico o cromoeléctrico en ciertas regiones del espacio. Tales variaciones de un campo cuántico se llaman fluctuaciones. En el campo electromagnético creado entre dos partículas cargadas eléctricamente, la presencia de fluctuaciones cuánticas implica que las interacciones entre las dos cargas no están estrictamente determinadas por el campo clásico predicho por las ecuaciones de Maxwell. Antes bien, el campo eléctrico medido es el promedio de todos los campos que pueden generarse por las fluctuaciones cuánticas, ponderados por la probabilidad de que ocurra cada fluctuación.

Para la mayoría de aplicaciones prácticas de la electrodinámica, los efectos producidos por las fluctuaciones cuánticas son muy pequeños. Si tomáramos una medición del campo entre dos objetos macroscópicos eléctricamente cargados, su valor no se alejaría del que ofreciera la teoría clásica. Sin embargo, en las colisiones a alta energía entre partículas cargadas, las fluctuaciones mecánico-cuánticas adquieren una significación más notable y deben tenerse en cuenta a la hora de calcular los efectos electromagnéticos. En el método convencional (que para muchos problemas es suficiente), primero se calcula las propiedades del campo en el vacío clásico. Sobre los resultados del cálculo clásico, se introducen a continuación las correcciones debidas a las fluctuaciones mecánico-cuánticas, de forma progresivamente compleja, según lo que se llama un desarrollo perturbativo. En la teoría cuántica del electromagnetismo, cuanto mayor o más complicada sea una fluctuación, menos probable será que ocurra. Por tanto, las correcciones al campo electromagnético clásico que deben realizarse en los cálculos mecánico-cuánticos obedecerán a fluctuaciones pequeñas.

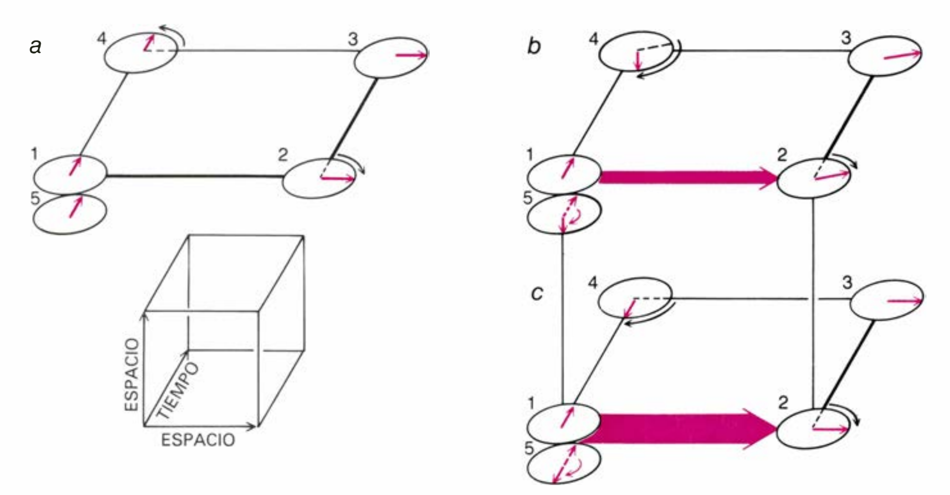
Alguien sospecharía que las propiedades del campo cromoeléctrico entre dos quarks podrían deducirse de manera análoga. Parecería, al menos en principio, que un desarrollo perturbativo proporcionara la intensidad del campo en cualquier punto con cualquier grado de aproximación necesario. Resulta, sin embargo, que el método del desarrollo perturbativo sólo funciona si el campo calculado para el vacío clásico es el efecto dominante. En otras palabras, el método sólo se muestra eficaz en el caso en que las correcciones a introducir en razón de las fluctuaciones son pequeñas y se hacen menores todavía a medida que se consideren fluctuaciones de tamaño creciente. Para los fenómenos mecánico-cuánticos que dependen principalmente de los efectos de grandes fluctuaciones, el desarrollo perturbativo no converge, es decir, las series de cálculos no tienden a un valor constante, finito. De tales fenómenos se dice que son no perturbativos. El confinamiento de las líneas de fuerza del campo cromoeléctrico entre dos quarks, y por tanto la ligadura permanente de los dos quarks, constituye un fenómeno no perturbativo.

¿Cómo se demuestra, pues, el confinamiento? Los físicos teóricos reconocieron que había que idear un nuevo enfoque en el que, desde un comienzo, se tomaran en consideración las grandes fluctuaciones del campo mecánico-cuántico. El método de la red sugerido por Wilson es una aproximación de éstas.

La red es generalmente cúbica y puede imaginarse como las aristas y vértices de una colección bien empaquetada de cubos. La red se extiende en el tiempo y en el espacio, es decir, cada punto de la red designa tanto una posición en el espacio como un instante en el tiempo. Para visualizar la red, piénsese en una serie de cubos donde dos ejes representen coordenadas espaciales y, el tercero, coordenadas temporales. La red entera tiene tres dimensiones espaciales y un eje de tiempo; se trata, pues, de una estructura cuadridimensional. Entre dos vértices vecinos de la red hay un enlace o arista que puede imaginarse como una línea que conecta los dos vértices. Un pequeño cuadrado limitado por cuatro aristas se llama plaquita (plaquette). En la formulación de Wilson los vértices, las aristas y las plaquitas de la red son todo lo que queda del espacio y el tiempo físico ordinario.

Aristas y plaquitas de la red deben considerarse como entidades de un nivel de abstracción distinto del que corresponde a los vértices o puntos de la red. Aunque las aristas y las plaquitas se definen por los puntos de la red, no hay puntos de la red adicionales a lo largo de una arista o dentro de una plaquita. En otras palabras, el espacio y el tiempo en la red es muy distinto del espacio y el tiempo ordinarios, que siempre incluyen un número infinito de puntos entre dos puntos dados.

La introducción por Wilson de una red espacio-temporal no implica que los procesos físicos se den en una red real. Según todos los datos disponibles, el espacio-tiempo es continuo. La red, por el contrario, representa lo que un físico teórico llama una regularización, un truco para hacer cálculos que de otra manera serían imposibles. Aplicada al problema del confinamiento, la



7. GENERALIZACION DE LA IDEA DE CAMPO DE AFORO. Basta con permitir que las cantidades definidas en cada punto de la red varíen en un intervalo continuo. Por ejemplo, déjese que el sentido de una flecha en cada punto forme un ángulo arbitrario con la vertical. Un campo de aforo permite comparar los ángulos en diferentes puntos de la red. Si una flecha no vuelve a su orientación original después del circuito completo de una plaquita, la diferencia angular entre las orientaciones inicial y final (una cantidad llamada ángulo de fase) mide el grado de frustración de la plaquita. El campo eléctrico es un tal campo de aforo, y la intensidad del campo en una dirección determinada en un punto cualquiera se mide por el grado de frustración de una plaquita asociada. La ilustración explica que la frustración de las plaquitas está asociada con una única componente del campo eléctrico, precisamente la componente espacial en la dirección izquierda-derecha. Estrictamente hablando, la red se extiende en cuatro dimensiones, las tres espaciales y el tiempo; a la hora de visualizarla sólo podemos asignarle dos coordenadas espaciales y una temporal. Cuando un medidor se transporta en sentido contrario a las agujas del reloj alrededor de una plaquita que incluye la dirección espacial izquierda-derecha y la dimensión temporal, el campo de aforo hace que la flecha del medidor gire. Si la flecha vuelve a su posición inicial después de que el medidor recorre todo el circuito de la plaquita, la intensidad del campo es cero en la dirección espacial y no hay línea de campo (*a*). Cuando el ángulo de fase aumenta, crece también la intensidad del campo a lo largo de la dirección espacial (*b*-*c*), lo que se representa por el creciente grosor de las líneas de campo (*flechas coloreadas*) a lo largo de las coordenadas espaciales. Desde el punto de vista matemático, la rotación de la flecha en el medidor puede identificarse con la rotación de una flecha en el plano de los números complejos, es decir, aquellos números que tienen una parte real y una parte imaginaria.

estrategia es la siguiente. Las partículas se definen sólo en los vértices de la red; la intensidad del campo, sólo a lo largo de las aristas de la red. (De hecho, lo que se define en cada vértice es la probabilidad de que una partícula se encuentre allí. Pero no se establece la probabilidad de encontrar una partícula entre dos vértices contiguos.) Cuando no hay partículas, la simetría de las fluctuaciones de la red simplifica drásticamente el cálculo del campo medio eléctrico o cromoeléctrico generado por las fluctuaciones fuertes. Los campos son cantidades vectoriales y tienen por tanto intensidad y orientación; la misma probabilidad hay de que apunten en una dirección a lo largo de una arista que lo hagan en la dirección opuesta. Por tanto, en el estado vacío, sin partículas, el valor medio del campo eléctrico o cromoeléctrico a través de la red será cero.

Un cálculo similar, aunque algo más elaborado, puede hacerse para las grandes fluctuaciones del campo cuando en la red se definen una única partícula y su antipartícula correspondiente. En promedio, las fluctuaciones del campo tornan a cancelarse, excepto a lo largo de las aristas de la red que forman la trayectoria más corta entre la partícula y la antipartícula. Los resultados no dependen de la clase de campo definido en la red; las partículas pueden ser un quark y su antiquark o un electrón y un positrón. Así, pues, el confinamiento será un resultado natural que se desprende de la definición del campo en la red.

La estrategia dicta que el siguiente paso a dar sea el de eliminar la red y recuperar el espacio-tiempo ordinario, continuo. El espaciado de la red se acorta de manera progresiva: los vértices se acercan más y se hacen más densos en el espacio-tiempo. Si la reducción del espaciado de la red continúa hasta el límite matemático, se recupera el espacio-tiempo continuo. En el límite, el procedimiento da el campo medio, atendidas todas las fluctuaciones mecánico-cuánticas.

¿Qué se ha ganado introduciendo la red? Desde un comienzo, y a lo largo de cada arista de la red, se tienen en cuenta las fuertes fluctuaciones, responsables del empaquetamiento de las líneas de fuerza de color. La estrategia impone un coste: mientras la red es relativamente poco tupida, las fluctuaciones cuánticas confinan el campo electromagnético, no sólo el cromoeléctrico.

A diferencia de las cargas de color, las cargas eléctricas existen aisladamente. Como el método de la red predice el confinamiento de la carga eléctrica, conviene mantenerse escéptico respecto al mismo, a menos que pueda demostrarse que a medida que la red se hace tupida el campo electromagnético se desenvuelve de acuerdo con sus bien conocidas propiedades. En otras palabras, lo que uno desearía demostrar es que, en algún momento del proceso de reducción del espacio de la red, las líneas de fuerza electromagnéticas se escapan de su confinamiento en una línea de aristas de la red, mientras que las líneas de fuerza cromoeléctricas permanecen comprimidas en todo el proceso de paso al límite del espacio-tiempo continuo. No es un asunto fácil demostrar que las cosas suceden precisamente de esta manera, pero en los últimos años la demostración llegó a través de elaborados cálculos numéricos, para los que se ha contado con la ayuda de computadores de alta velocidad.

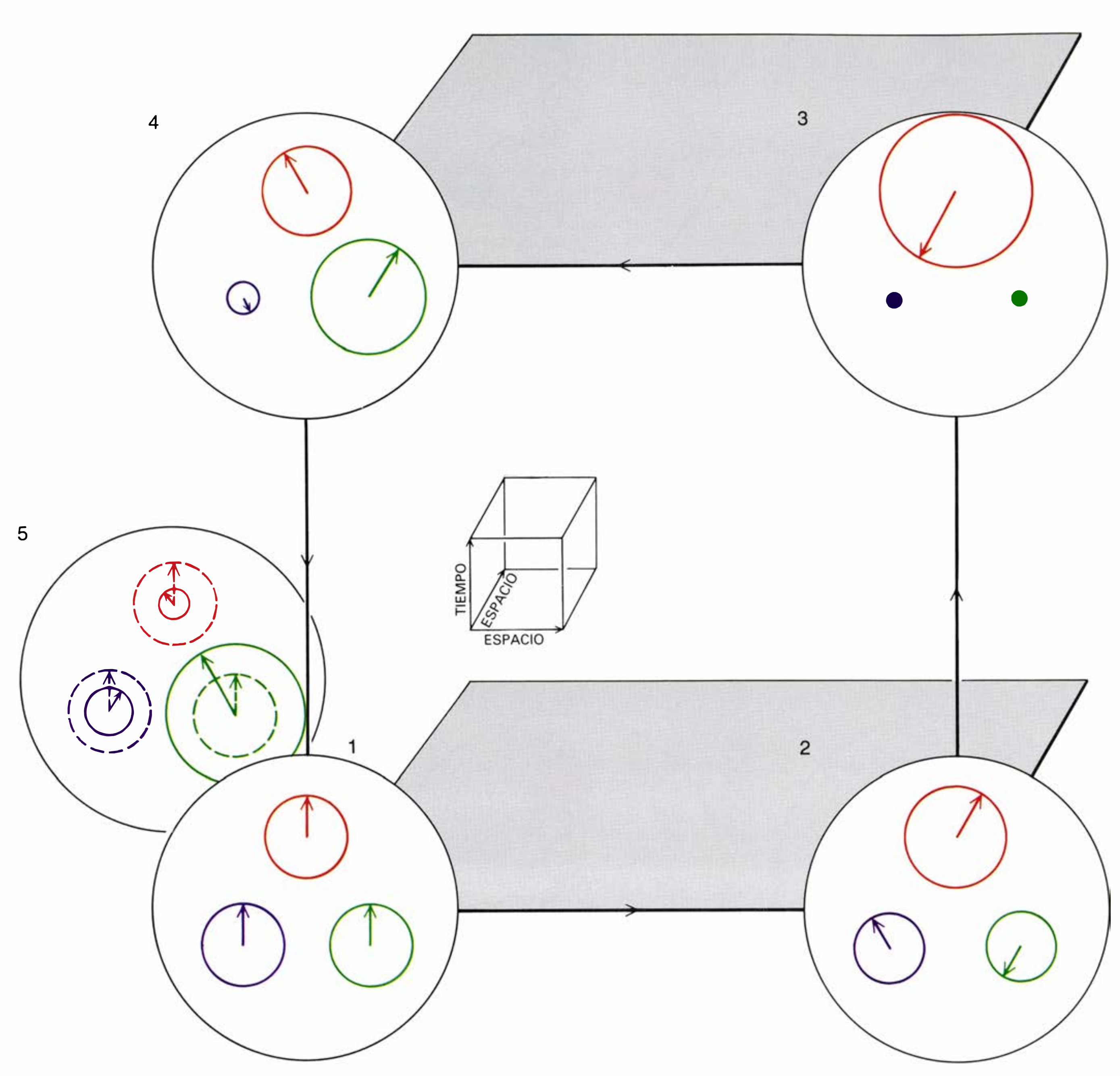
Así distinguen los físicos teóricos el campo electromagnético del campo cromoeléctrico: la QED es una teoría de aforo abeliana, mientras que la QCD es una teoría de aforo no abeliana. El adjetivo que marca la diferencia alude al matemático noruego Niels Henrik Abel. La distinción entre abeliano y no abeliano se debe a la teoría matemática de grupos que describe las simetrías inherentes a una secuencia de operaciones, una secuencia de rotaciones, por ejemplo. Si las operaciones que son miembros de un grupo pueden realizarse en cualquier orden con el mismo resultado final, el grupo es abeliano. Por ejemplo, el grupo de rotaciones en torno a un eje único es abeliano, porque tales operaciones tienen el mismo efecto con independencia de su orden. Por otro lado, si el orden en que se realizan dos o más operaciones afecta al resultado final, el grupo de operaciones se llama no abeliano. Las rotaciones de un cubo alrededor de sus tres ejes forman un grupo no abeliano: cuando el cubo se gira alrededor de un eje vertical y uno horizontal, el resultado depende de qué operación se ha hecho primero.

Para entender cómo se aplica la teoría de grupos a la QCD y la QED, conviene conocer primero el concepto de campo de aforo. La idea se ilustra mejor para los puntos aislados del espacio y el tiempo ordenados en una red. Las partículas pueden asentarse en algún vértice de la red o saltar de un vértice a otro; cuando las partículas se

mueven a través de la red espaciotemporal pueden cambiar su estado, entendiendo por estado unas cantidades que varían en cierto rango de valores. Partamos de una hipótesis simplificadora: establezcamos que el estado de una partícula se describa por una variable que pueda tomar sólo dos valores. En cada vértice de la red habrá una variable cuyo valor será +1 o −1, que expresa el signo de la carga eléctrica. Los dos posibles valores pueden representarse por una flecha que apunte hacia arriba o hacia abajo.

A la hora de exponer de qué forma las interacciones se propagan en el espacio y el tiempo necesitamos comparar los valores de las variables en puntos vecinos. Para comparar las longitudes de dos objetos en regiones distintas del espacio, se precisa una vara para medir, o aforo, que pueda situarse al lado de un objeto, marcarse y trasladarse al segundo objeto, con el que se hará la comparación. Análogamente, en la red uno ha de poder comparar la orientación de las flechas en vértices vecinos. Al principio una tal comparación parece trivial. Supongamos que dos vértices de la red se representan sobre una hoja de papel y en cada vértice haya una flecha dirigida hacia arriba. ¿No es obvio que las dos flechas apuntan en la misma dirección? La cuestión, sin embargo, presupone la existencia de la hoja de papel en la que se ha dibujado la red. El papel actúa de intermediario que permite comparar las dos orientaciones. En cierto sentido, uno transporta la flecha de un vértice al otro con el ojo, y concluye que las dos flechas tenían la misma orientación antes del transporte, porque después del mismo coinciden.

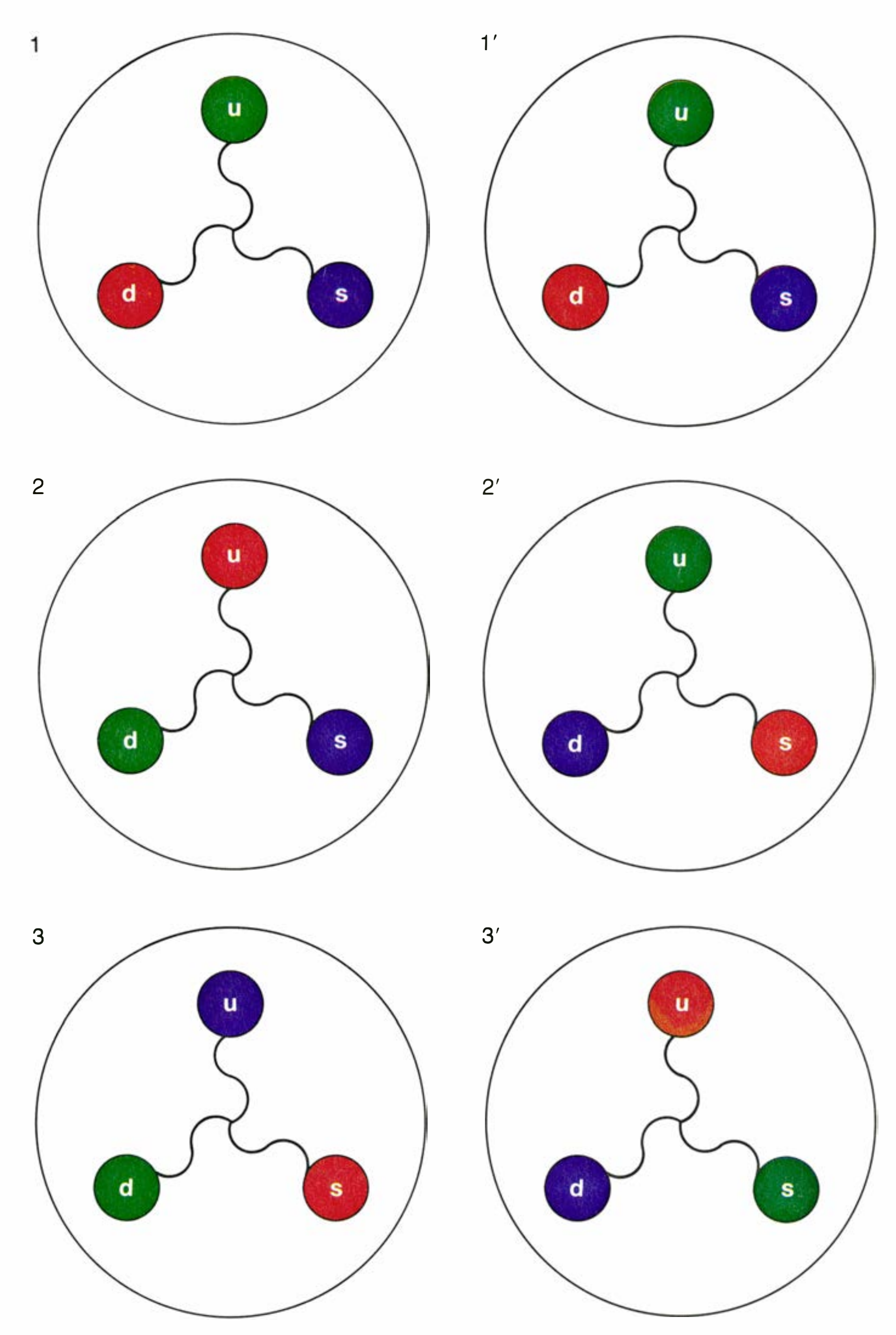
Imaginemos que se hubiera dado medio giro a la hoja de papel entre las dos flechas [*véase la figura 5*]. Una flecha hacia arriba transportada sobre el papel



8. ES EL CAMPO CROMOELECTRICO un campo de aforo similar en principio al campo electromagnético, aunque rodeado de una mayor complejidad matemática. En cada punto de una red hay tres flechas en vez de una; corresponden a las cargas de color de un quark. Además, el campo de aforo de color afecta no sólo a la dirección de cada flecha, sino también a su longitud. Las longitudes de las flechas no son independientes entre sí: la raíz cuadrada de la suma de los cuadrados de las longitudes debe ser igual a 1. La intensidad del campo cromoeléctrico a lo largo de cada arista de la red depende de los ángulos de fase y del cambio en la configuración de las tres flechas después de un circuito completo de una plaquita. (Ilustración de I. Arbel.)

torcido apuntaría hacia abajo al llegar al segundo vértice. Como la dirección de la flecha se define sólo en los vértices individuales, aislados, de la red, no hay manera de decidir cuál de los dos métodos de transporte es el correcto. Sin la hoja de papel o alguna hipótesis similar sobre los efectos del transporte no es posible una comparación de las direcciones en vértices diferentes.

En la teoría de campos, un aforo o contraste es cualquier patrón de medida, análogo a la distancia entre dos marcas en una barra metálica o la dirección de una flecha con respecto a una dirección de referencia, que puede cambiar bajo la influencia de un campo cuando el aforo se mueve en el espacio y el tiempo. Un campo que pueda inducir tales cambios se llama un campo de aforo; especifica explícitamente las hipótesis a asumir sobre el transporte del aforo. En el ejemplo dado aquí, el campo de aforo es un conjunto de reglas para transportar las flechas a lo largo de las aristas de la red de un vértice al siguiente. Uno puede imaginar el campo de aforo como una cinta que conecta puntos vecinos de la red y que permite, por tanto, transportar las flechas a puntos diferentes para compararlas.



9. GRUPO DE LAS TRANSFORMACIONES DE COLOR en el campo cromoeléctrico. Se clasifica en matemática como no abeliano. Un grupo no abeliano es un conjunto abstracto de operaciones (junto con los objetos sobre los cuales se realizan las operaciones) en que el resultado final de varias operaciones depende del orden en que se realicen. En la teoría de aforo cromoeléctrica, las operaciones son intercambios de color entre los quarks. En esta ilustración se aplican dos intercambios de color en una configuración de quark dada (*1a*, *1b*). En la izquierda, un cambio de rojo y verde (*2a*) es seguido por un cambio de violeta y rojo (*3a*). En la fila de la derecha, el orden de los dos intercambios se invierte (*2b*, *3b*).

Importa advertir que la flecha que representa el estado en cualquier vértice de la red puede invertirse, siempre que la cinta que representa el campo de aforo se tuerza o no, según corresponda. Si las flechas y las cintas se ajustan coordinadamente, el resultado de cualquier comparación entre las variables de estado no cambia, como tampoco lo hace la información física representada por el sistema. La posibilidad de invertir las flechas sin modificar la información física se llama invariancia de aforo local.

Con la invariancia de aforo local por delante, diríase que el campo de aforo es una complicación innecesaria. ¿Qué interés esconde introducir la cinta torcida, si el giro se puede deshacer por una transformación de aforo local? La objeción es válida cuando sólo consideramos pares de puntos de la red. Sin embargo, la importancia de la invariancia de aforo local se hace patente en cuanto miramos hacia una plaquita de la red: un conjunto de cuatro puntos de un cuadrado, conectados, a lo largo de las aristas de la red, por cuatro cintas que constituyen el campo de aforo. Supongamos que se tuerce una de las cintas y las otras tres no [*véase la figura 6*]. La única cinta torcida puede deshacer su giro (e invertir una de las flechas), pero a su vez se introducirá un semigiro en una de las otras tres cintas. Con independencia de cuál sea la flecha que se invierta y de cuántas transformaciones de aforo locales se hagan, se torcerá un número impar de las cuatro cintas. De la plaquita cuyas cintas no puedan estar todas enderezadas se dice que está frustrada.

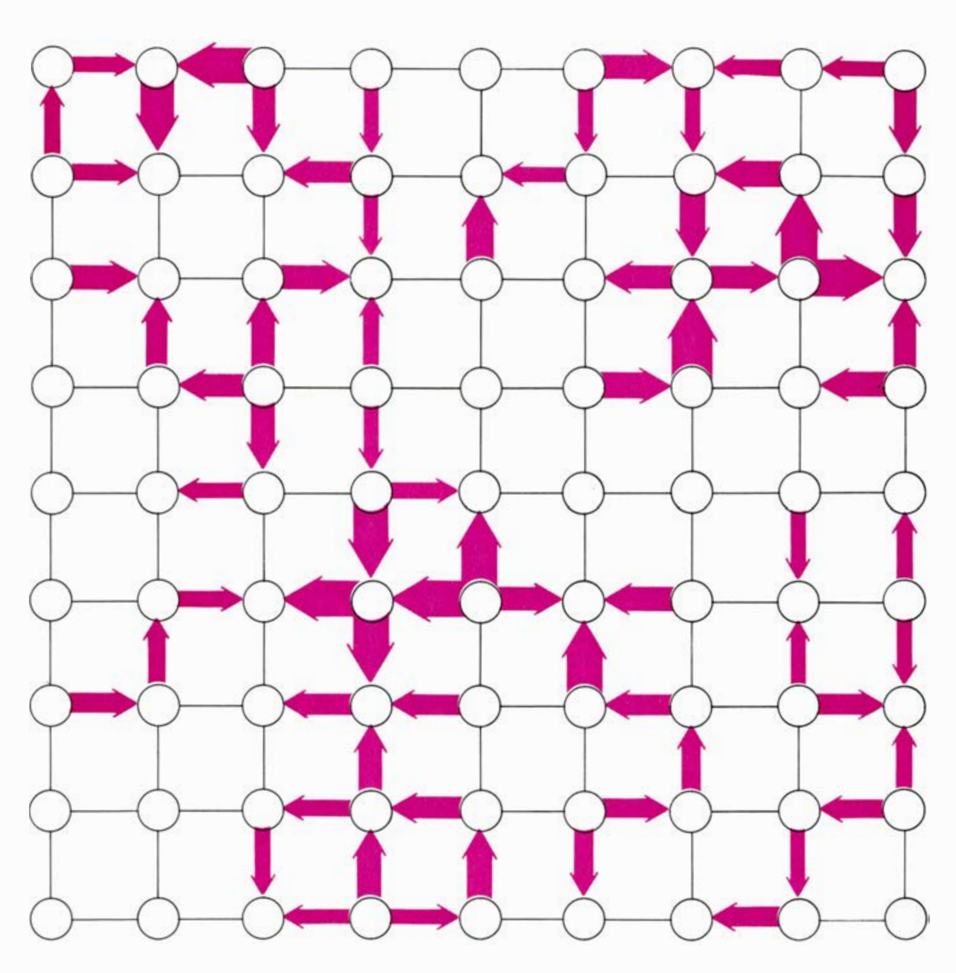
La frustración de la plaquita se puede manifestar de otra manera. Una flecha que apunte hacia arriba, que se transporte a lo largo de las cintas alrededor de la plaquita frustrada, volverá a su vértice de salida apuntando hacia abajo. Así, las plaquitas frustradas de la red son aquellas que presentan un desajuste en la orientación de las flechas cuando se transportan en todo su perímetro.

No es difícil aplicar la idea de una plaquita frustrada al espacio-tiempo ordinario. En el espacio-tiempo físico, las cintas a lo largo de las cuales se mueven las flechas no pueden visualizarse directamente; igual que las hojas de papel que definen el campo de aforo en la red no forman parte de la red, así un campo

de aforo definido en el espacio-tiempo ordinario no es él mismo parte del espacio-tiempo. Matemáticamente, el espacio abstracto de más dimensiones que especifica las rotaciones de la flecha se conoce por conexión de fibrado [véase "Fibrados y teoría cuántica", por Herbert J. Bernstein y Anthony V. Phillips; INVESTIGACIÓN Y CIENCIA, septiembre, 1981]. A pesar de todo, no se necesita dominar toda la teoría matemática subyacente para imaginar que una flecha que se mueve a lo largo de una trayectoria cerrada en el espacio-tiempo pueda volver a su punto inicial con su dirección cambiada.

La idea de plaquita frustrada tiene una interpretación física importante. En cualquier campo de aforo, la energía del campo reside en esas plaquitas frustradas. Las no frustradas, en las que todas las cintas pueden estar enderezadas y todas las flechas orientarse en la misma dirección, se asocian con el estado vacío de un sistema físico, la configuración carente de energía. Las plaquitas frustradas son las sedes de las fluctuaciones en un campo mecánico-cuántico.

La idea de un campo de aforo puede generalizarse a situaciones más complicadas: permitir, por ejemplo, que las flechas formen un ángulo arbitrario con respecto a una dirección fija, en vez de constreñirlas a apuntar hacia arriba o hacia abajo. El campo de aforo especifica ahora el ángulo que gira la flecha cuando se transporta a lo largo de una arista. Consiguientemente, la frustración de una plaquita puede variar en un rango de valores continuos, medidos por la diferencia angular en la dirección de la flecha una vez transportada alrededor de toda la plaquita. El grado de frustración de una plaquita, expresado en unidades matemáticas adecuadas, se llama su acción. La acción del sistema es la suma de las acciones de las distintas plaquitas.

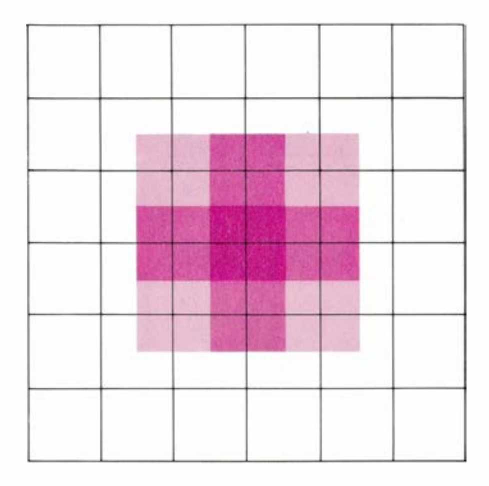
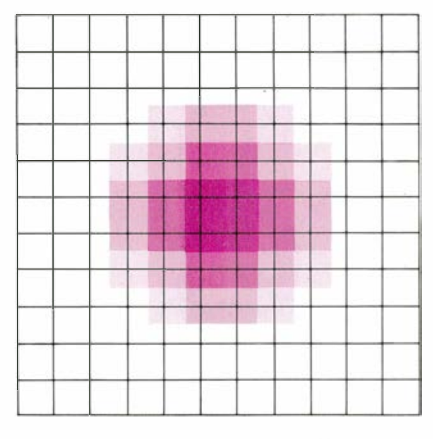


10. FLUCTUACIONES de las líneas de fuerza en una red, características del vacío mecánico-cuántico. Las fluctuaciones son consecuencia del principio de incertidumbre de Werner Heisenberg. El principio establece que el producto de la incertidumbre sobre la energía de un sistema y la incertidumbre acerca del tiempo de un suceso no es menor que una constante numérica. Por tanto, durante cualquier intervalo corto hay una incertidumbre correspondientemente grande en la energía del sistema, que se manifiesta en fluctuaciones espontáneas del campo cuántico. Las fluctuaciones se cancelan entre sí en ausencia de partículas cargadas; una medida del campo eléctrico en el vacío daría el valor cero porque determinaría sólo el promedio de todas las fluctuaciones. Pero las fluctuaciones de las líneas de fuerza en el vacío deben considerarse con sumo cuidado al calcular el campo de fuerzas entre dos partículas cargadas.

Resulta así que el eléctrico es un campo de aforo que especifica las rotaciones continuas de una flecha en torno a un eje fijo, aunque este hecho sólo se explicite en las formulaciones más unificadas y refinadas del concepto de campo eléctrico. Lo corriente es pensar en el campo eléctrico como un conjunto de vectores, con un vector en cada punto del espacio que indica la magnitud y dirección del campo en aquel punto. Sin embargo, tal como se entiende en el enfoque más lato de la teoría de aforo, la magnitud de cada vector es directamente proporcional a la frustración de una plaquita asociada.

La flecha giratoria del campo de aforo asociada con el electromagnetismo representa un número complejo (dícese del que tiene una parte real y una parte imaginaria), que cambia de valor cuando se transporta la flecha alrededor de la plaquita. (No hemos de confundir la flecha transportada y el número complejo con el vector que representa el propio campo eléctrico en cada arista de la red.) La plaquita es un bucle en el espacio-tiempo; y es, por tanto, necesario imaginar la flecha transportada en movimiento hacia delante y hacia atrás, en el tiempo y en el espacio. La flecha se transporta a lo largo de cierto eje en el espacio, por ejemplo el eje x positivo, después hacia adelante en el tiempo, regresa luego a su punto de partida en el eje x y, finalmente, se retrasa hasta su punto inicial en el tiempo. La cantidad de cambio direccional de la flecha transportada resultante del transporte se llama el ángulo de fase. Según la teoría de aforo, la magnitud de la componente x del vector que ordinariamente da la intensidad del campo eléctrico en un punto es una medida del ángulo de fase generado alrededor de un bucle espacio-temporal que empieza en la dirección x. Las fluctuaciones mecánico-cuánticas del campo eléctrico en un punto pueden imaginarse, por tanto, como fluctuaciones en la cantidad de rotación que una flecha sufriría si fuera transportada alrededor de una plaquita extendida a lo largo de una dimensión espacial y a lo largo de la dimensión temporal [*véase la figura 7*].

Es difícil evitar la sensación de irrealidad cuando se nos sugiere considerar el campo eléctrico –algo que puede resultar bastante tangible si sufrimos una descarga– como el espacio abstracto de las rotaciones de fase. Pero a medida que vamos profundizando en el estudio del mundo físico, los hechos tangibles y los conceptos matemáticos se entrelazan. La idea abstracta del espacio de fase termina por establecer contacto con la distinción entre la QED y la QCD, es decir, con la distinción en-

11. REDUCCION DE UNA RED en la que se han definido las partículas y los campos. Debe realizarse al objeto de aproximar los efectos del campo en el espacio-tiempo ordinario, continuo. A medida que el tramado o espaciado de la red se estrecha, el tamaño de una partícula compuesta, piénsese en el protón, deberá mantenerse; significa ello que la probabilidad de que un quark se encuentre en un vértice dado de la red dentro del protón se extenderá sobre un número creciente de vértices. La ilustración de la derecha muestra el resultado de reducir la red de la izquierda hasta la mitad de su tamaño anterior. La densidad de color representa la probabilidd relativa de que un quark ocupe un vértice dado. A semejante reducción se llega variando de un modo sistemático el valor de la cantidad conocida como constante de acoplamiento, que determina la intensidad de las fluctuaciones en el campo mecánico-cuántico.

tre una teoría de aforo abeliana y otra no abeliana. Las operaciones de grupo fundamentales en la QED son las rotaciones de ángulos de fase. El resultado de dos rotaciones de fase sucesivas, digamos una de 30 grados y otra de 50 grados, no depende del orden en que se hagan. En cualquier caso, el resultado es una rotación de 80 grados. Tal resultado es característico de un grupo abeliano y, por tanto, la QED es una teoría de aforo abeliana.

Si las variables de estado definidas en los vértices de una red son los colores de los quarks, el aforo transportado está formado por tres flechas, y no sólo por una; las flechas variarán, además, en longitud y en ángulo [*véase la figura 8*]. Cada flecha representa una de las tres cargas de color; el conjunto de las cargas de color determina el color del quark. Cuando el quark se transporta a lo largo de una arista de la red puede cambiar su color de rojo a verde; por tanto, un quark rojo puede intercambiar colores con un quark verde, y el nuevo quark verde hará entonces lo propio con un quark violeta. Sin embargo, el resultado final de los intercambios depende del orden de los sucesos. Ahora bien, un resultado que depende del orden es característico de un grupo no abeliano; por consiguiente, la QCD será una teoría de aforo no abeliana.

La naturaleza no abeliana de la QCD introduce un grado de libertad adicional en las fluctuaciones del campo cromoeléctrico. Además, la existencia de tres clases de color implica también que las plaquitas del campo cromoeléctrico puedan frustrarse de muchos más modos que las plaquitas del campo electromagnético. Es probable que el grado de libertad extra y el desorden adicional de la teoría no abeliana sean responsables del confinamiento de los quarks dentro de los hadrones.

Ya he dicho que el confinamiento en una red poco tupida puede demostrarse mediante un cálculo bastante sencillo de las grandes fluctuaciones del campo mecánico-cuántico. Sin embargo, para fluctuaciones menores en una red más fina, el cálculo del campo adquiere una complejidad mayor. El valor de cualquier cantidad física medida en un experimento es el valor esperado cuántico, que es un promedio ponderado de todos los posibles valores que la cantidad tome. El campo eléctrico o cromoeléctrico medido es el promedio de todas las posibles configuraciones del campo que las fluctuaciones originen. No todas las configuraciones contribuyen por igual a la media; cada configuración debe ser ponderada, o multiplicada por algún factor relacionado con la probabilidad de la configuración. Por tanto, el valor esperado cuántico de una cantidad física definida en la red puede calcularse, en principio, en dos pasos. Se determina primero el valor para cada configuración de las fluctuaciones del campo y se multiplica por el factor de peso de dicha configuración. Después se suman los productos y el resultado se divide por la suma de los factores de peso.

Incluso para una red definida en un pequeño volumen del espacio-tiempo, el número de configuraciones posibles es tan grande que resulta impensable una suma completa. Para la teoría de aforo más simple posible, en la que el campo de aforo entre dos vértices de la red puede estar o no torcido, el número de configuraciones en una red de 10 vértices en cada dirección es $2^{40.000}$, es decir, más de $10^{12.041}$.

Para calcular el valor cuántico esperado del campo en la red se debe acudir a técnicas de muestreo estadístico. Las técnicas son análogas a las empleadas en las encuestas de opinión. No podemos preguntar la opinión a todos los ciudadanos de un país para determinar qué piensan sobre algo distintos grupos de población; se selecciona una muestra. El resultado debe reflejar la verdadera opinión de los grupos si la probabilidad de seleccionar un encuestado de cierto grupo coincide con la fracción, o peso, del grupo en el conjunto de la población.

De manera similar, las configuraciones de las fluctuaciones del campo se eligen por un programa de computador. El computador genera un número elevado de configuraciones (inferior, por supuesto, a $10^{12.041}$), y la probabilidad con que se genera una configuración determinada se iguala al factor de peso mecánico-cuántico de dicha configuración. La muestra de configuraciones tiende a dar el mismo valor medio del campo cuántico que daría todo el conjunto de configuraciones.

La importancia de una configuración en el cálculo del valor esperado cuántico se determina por su acción, que se designa en general por S; por tanto, el factor peso está dado por una función matemática de la acción S. Una fórmula elegante para el factor de peso en los sistemas mecánico-cuánticos fue introducida por Richard P. Feynman, del Instituto de Tecnología de California. Primero, se divide la acción total de la configuración por una constate, g^2; después, se determina la exponencial de esta cantidad; en otras palabras, el número e (igual aproximadamente a 2,7) se eleva a la potencia S/g^2. El factor de peso es inversamente proporcional al resultado: el factor de peso es proporcional a $1/\exp(S/g^2)$.

La fórmula del factor de peso implica que, cuanto más alta sea la acción de una configuración, menos pesará la configuración en el cálculo del campo cuántico medio. No exponencia la propia acción, sino que se divide por g^2; por consiguiente, el cambio del valor de g puede tener un efecto significativo en el factor de peso de una configuración dada. Con g grande, S/g^2 será pequeño y $1/\exp(S/g^2)$ mayor que lo que sería si g fuera pequeño. La cantidad g

se llama constante de acoplamiento. Ante una constante de acoplamiento elevada, el factor de peso de una fluctuación mecánico-cuántica con gran acción es más alto que cuando la constante de acoplamiento es pequeña.

La idea de una constante de acoplamiento puede resultarnos familiar cuando se trata de medir la intensidad intrínseca de una fuerza. En la teoría electromagnética, la constante de acoplamiento es una cantidad física importante, con un valor cercano a 1/137. Desde la perspectiva más abstracta de la teoría de aforo en una red, g representará una cantidad que toma un valor fijo para un espaciado de la red, aunque varíe con el espaciado. Una vez demostrado el confinamiento en una red poco tupida, el entramado de ésta se hace más fino y disminuye cuidadosamente el valor de la constante de acoplamiento. El proceso por el que la red adquiere progresiva sutileza hasta que se recupera el espacio-tiempo continuo recibe el nombre de renormalización.

Hablando con entera propiedad, no resulta posible todavía renormalizar la red para fluctuaciones del campo cromoeléctrico, o en otras palabras, reducir el espaciado de la red a cero. Sí podemos acortar el espaciado de la red manteniéndolo mayor que cero y buscar indicaciones que nos manifiesten si las líneas de fuerza permanecerán o no confinadas en redes más finas. A medida que el espaciado de la red se restrinja, todos los objetos físicos permanecerán del mismo tamaño. En una red poco tupida, la distribución de probabilidad para un protón se define como no nula sólo a lo largo de tres espaciados de la red. Cuando el espaciado queda a la mitad de su valor original, la distribución de probabilidad debe extenderse sobre seis espacios de la red. Como la reducción del valor de g baja la probabilidad de las configuraciones de la red que tienen acción grande, la disminución de g permite una mejor inspección de las fluctuaciones del campo, de las líneas de fuerza y de las partículas definidas en la red. Al reducir el valor de g se logra que el protón parezca mayor en la red.

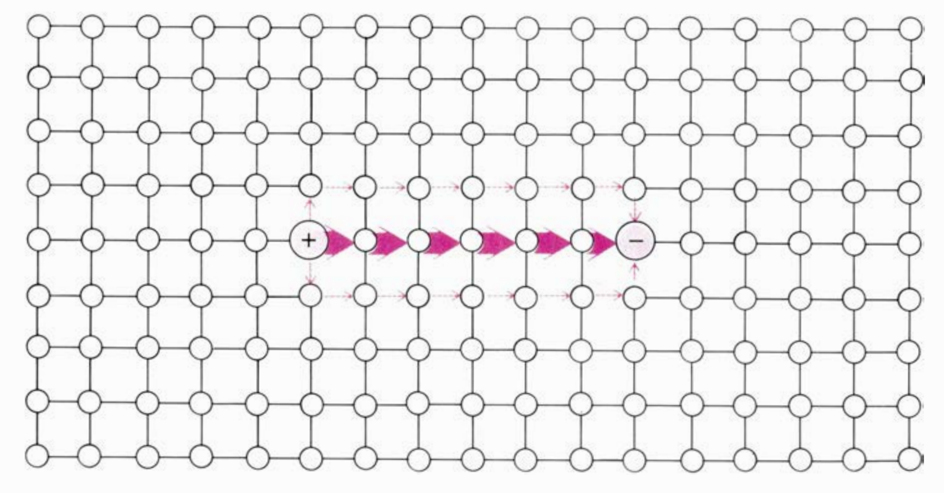
La investigación numérica de las propiedades de un campo de aforo en una red procede limitando ésta a un volumen grande, aunque finito. El número de vértices, aristas, plaquitas y variables de estado de la red será, pues, finito, aunque supere los 100.000. Los valores iniciales de todas las variables se almacenan en la memoria de un potente ordenador. Variando al azar los elementos de la configuración inicial según un algoritmo adecuado, el computador genera una muestra de hasta 100.000 configuraciones. Y calcula luego los efectos mecánico-cuánticos medios producidos por las configuraciones de la muestra.

Los elementos aleatorios del cálculo dan nombre al método: simulación de Monte Carlo. Antes de que mis colegas en el tema del confinamiento de los quarks y yo mismo hubiéramos recurrido a él, el método de Monte Carlo se había aplicado con considerable éxito al análisis de las propiedades de los sistemas termodinámicos. El propio Wilson había resaltado su idoneidad para el análisis de teorías de aforo en una red en mecánica cuántica. En 1979, trabajando en Brookhaven, Michael J. Creutz, Laurence A. Jacobs y el autor aplicaron por primera vez la simulación de Monte Carlo al estudio de las teorías de aforo abelianas. Pretendíamos probar si el confinamiento de las partículas en la red observado para grandes valores de la constante de acoplamiento en la QED desaparecía, lo que había de ocurrir siempre que se aproximara al límite continuo. Los resultados fueron espectaculares. Mostraron con nitidez que, en algún paso a través de la reducción de la constante de acoplamiento, las líneas de fuerza de la red sufrían de repente una transición. El campo eléctrico, que para grandes valores de g estaba confinado a las aristas de la red entre dos cargas eléctricas, se extendía de repente a todo el entorno de las cargas.

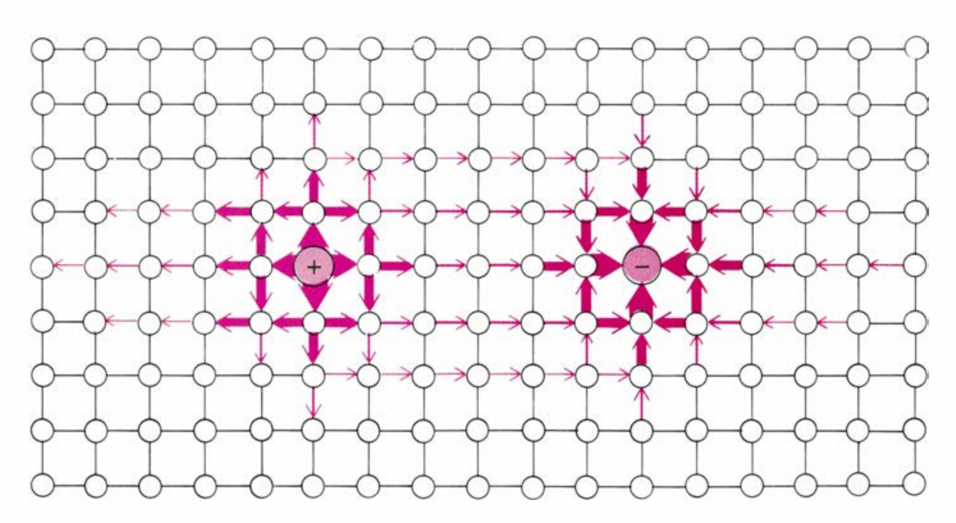
Compárese el súbito desconfinamiento del campo eléctrico con el cambio repentino de las propiedades observadas cuando el sólido pasa a fase líquida. Creutz, Jacobs y yo observamos en nuestros cálculos numéricos una transición de fase. En el modelo computacional todo sucedía como si tuviéramos un cristal cuadridimensional que podíamos calentar o enfriar modificando el valor de g. En cierto punto, el cristal sufría una transición de fase; en un lado de la transición las cargas eléctricas están confinadas, mientras que en el otro lado no lo están. Más tarde, nuestros resultados recibían ulterior, e independiente, confirmación de una extensa simulación numérica realizada en la Organización Europea para la Investigación Nuclear (CERN), por Benny Lautrup, del Instituto Niels Bohr, y Michael Nauenberg, de la Universidad de California en Santa Bárbara, y otros investigadores.

Poco después de que obtuviéramos nuestros resultados para la teoría de aforo abeliana en una red, Creutz generalizó la simulación de Monte Carlo al modelo no abeliano. Sus resultados fueron tan espectaculares, y más interesantes desde el punto de vista físico, que los nuestros. Con la QED se conocía de antemano la respuesta correcta y la simulación constituía más una prueba del método de Monte Carlo que de la QED; pero el resultado para el sistema no abeliano era totalmente desconocido. La simulación de Creutz mostró que, al contrario del caso abeliano, la teoría de aforo no abeliana en una red no presentaba una transición de fase cuando el valor de la constante de acoplamiento se reducía gradualmente. Así, se corrió el velo de algo que se había largamente sospechado: se probó, aunque por métodos numéricos, que la cromodinámica cuántica conducía al confinamiento de los quarks.

La simulación de Monte Carlo permite explorar las predicciones de la QCD para muchos procesos físicos.



12. CONFINAMIENTO DE UN CAMPO, ya sea el campo eléctrico o el cromoeléctrico; puede demostrarse por cálculos relativamente sencillos cuando el campo se define en una red ancha y las fluctuaciones de las líneas de fuerza son intensas. Las fluctuaciones intensas se cancelan entre sí, en todas las partes de la red excepto en la región que media entre las dos partículas cargadas. Las líneas de fuerza se comprimen dentro de un tubo estrecho; de ese modo, las dos partículas no pueden alejarse indefinidamente.



13. CALCULO DEL CAMPO ELECTRICO generado por dos partículas de cargas opuestas definidas en una red, realizado por el autor con un computador Control Data Corporation CDC-7600 en el Brookhaven National Laboratory. La ilustración muestra el resultado de miles de cálculos, realizados para intervalos de la red progresivamente menores. Cuando el intervalo se reduce, las líneas de fuerza que están confinadas en una red ancha sufren de repente una transición de fase y se esparcen en todas direcciones. El levantamiento del confinamiento del campo eléctrico en la red confirma la plausibilidad del método, porque es bien sabido que el campo eléctrico se esparce en el espacio ordinario, continuo. Cálculos similares realizados por Michael J. Creutz han demostrado que, cuando el intervalo de la red se reduce para el campo cromoeléctrico, las líneas de fuerza permanecen confinadas, tal como lo están en una red muy suelta. El confinamiento de los quarks puede derivarse de la teoría del campo cromoeléctrico.

Por ejemplo, después de que Creutz mostrase que no hay una transición de fase desconfinante en la QCD, pudo también estimar la fuerza que mantenía unidos a los quarks. El resultado está en excelente acuerdo con las predicciones del modelo de la cuerda, que no pretende describir la dinámica de los quarks y los hadrones en toda su generalidad. Confirmaron el valor de Creutz para la tensión de la cuerda Wilson, Gyan Bhanot, del CERN, el autor y otros.

Predice la simulación de Monte Carlo este otro resultado sorprendente: a temperaturas extremadamente altas, se levanta el confinamiento a los quarks, que pueden moverse libremente. Las temperaturas deben ser cercanas a dos billones de grados Celsius, subidísimas para las posibilidades del laboratorio, pero quizá no demasiado altas para que haya quarks libres en el interior de estrellas calientes. La predicción sugiere que existieron quarks libres inmediatamente después de la gran explosión.

La frontera actual que tiene ante sí nuestro conocimiento de las interacciones de los quarks es la simulación numérica de los cambios en el campo cromoeléctrico a medida que los quarks se desplazan. Todas las investigaciones descritas hasta ahora abordan el campo cromoeléctrico suponiendo que los quarks son estacionarios. Nos es permitido emplear el método de Monte Carlo para simular la mecánica-cuántica de los quarks en movimiento en un campo de aforo, pero la cantidad de cálculo necesario coloca la simulación casi fuera de alcance de los computadores más potentes. Herbert W. Hamber, del Instituto de Estudios Avanzados de Princeton, Enzo Marinari y Georgio Parisi, de la Universidad de Roma, y el autor, e independientemente Donald H. Weingarten, de la Universidad de Indiana, hemos propuesto un guión teórico en cuyo marco los cálculos son factibles. La aproximación se ha aplicado al cálculo de las masas de varios hadrones, y los resultados teóricos concuerdan bastante con los valores experimentales.

La teoría de aforo en una red ha catapultado por fin a la QCD hasta la fase en que podemos ya calcular sus predicciones y compararlas con el experimento. Es una teoría elegante basada en conceptos matemáticos relativamente sencillos y rica en consecuencias. Hasta el momento, todas sus predicciones verificables están superando la prueba experimental. Resta por lograr que sus consecuencias se prueben por deducción lógica. Aunque una aproximación numérica no puede satisfacer la necesidad de una demostración lógica, es un buen augurio para su futura demostración. Abundan los ejemplos, en otras ramas de la física teórica, según los cuales un resultado que se obtuvo primero numéricamente, no tardó en corroborarse por métodos analíticos. El autor comparte la opinión de algunos colegas suyos: la QCD pasará también pronto esta prueba.

Vacunas sintéticas

Una breve cadena de aminoácidos, sintetizada in vitro cual segmento de la superficie de una proteína vírica, induce anticuerpos que confieren inmunidad ante el virus

Richard A. Lerner

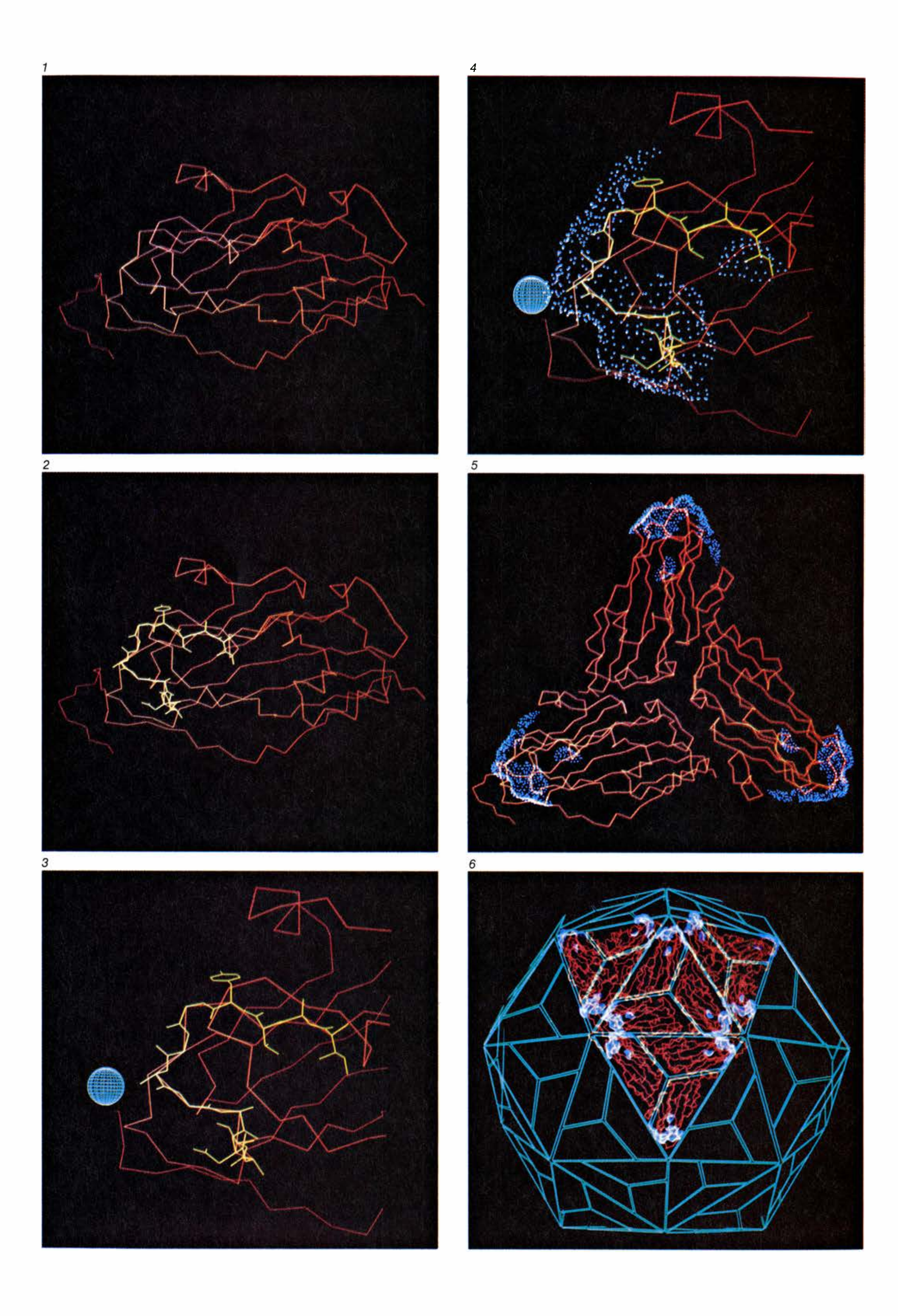
La vacuna es uno de los descubrimientos de mayor alcance de cuantos ha realizado el hombre. Ha erradicado la viruela y ha permitido controlar enfermedades como la difteria, la poliomielitis o el sarampión. A pesar de lo cual, no podemos decir que el procedimiento sea perfecto del todo: la seguridad de una vacuna no puede garantizarse absolutamente. La inmunización contra las infecciones víricas se obtiene por inyección de un virus que estimula la producción de anticuerpos capaces de neutralizar al agente patógeno; el virus inyectado no causa infección, puesto que se ha atenuado o matado previamente. Un virus atenuado es aquel cuya virulencia en humanos ha disminuido, por mutación, durante su paso por algún huésped animal; en ese proceso, el virus queda mejor adaptado a vivir en ese animal que en los seres humanos. (El virus de la viruela de las vacas, que induce inmunidad frente al virus de la viruela humana, es, de hecho, un virus atenuado naturalmente.) No deja por ello de ser un organismo vivo, capaz de cambiar. Puede seguir mutando, lo que acentuará, presumiblemente, su virulencia. Incluso en ausencia de mutación, el virus podría tener algún efecto desconocido a largo plazo, comparable con el de los virus "latentes", de acción lenta. Estas posibilidades desaparecen si se inactiva el virus; aún así, casos ha habido en que un virus insuficientemente inactivado ha producido la enfermedad.

Los virus atenuados comparten ciertos problemas con los inactivados. Toda instalación donde se cultiven virus para la producción de vacunas está expuesta a que un accidente provoque la diseminación de agentes patógenos. Además, puesto que los virus sólo se replican en un sistema vivo, suelen mantenerse en cultivos celulares, en huevos embrionados, o se aíslan de la sangre de animales infectados (el virus de la hepatitis *B*, sólo de personas infectadas); las células en cultivo, los huevos, o la sangre pueden contener sustancias que pasen desapercibidas, especialmente otros virus, y contaminar la vacuna. Algunas vacunas, incluso las constituidas por virus inactivados, deben conservarse bajo refrigeración desde el momento en que se preparan hasta que se inoculan; mantener esta "cadena de frío" puede convertirse en un problema delicado en regiones subdesarrolladas.

De ahí el justificado interés creciente por la producción de vacunas sintéticas: vacunas que no contengan virus intactos, sino simples péptidos (breves cadenas proteicas) ensamblados en el laboratorio, que mimeticen una región muy pequeña de la envoltura externa del virus y que induzcan la producción de anticuerpos capaces de neutralizar el virus. En muchos laboratorios se han desarrollado ya vacunas experimentales de este tipo; también en el nuestro, del Instituto de Investigaciones de la Clínica Scripps. Ello ha permitido avanzar bastante en el conocimiento de la estructura inmunológica de las proteínas. La capacidad para sintetizar péptidos que induzcan anticuerpos de especificidad precisa y prestablecida ofrecerá a los biólogos moleculares una serie de herramientas cuya importancia competirá con el interés clínico de las vacunas sintéticas.

Antes de exponer nuestros experimentos y resultados, y lo que nos han revelado sobre la estructura de las proteínas, describiré cómo nos protege el sistema inmune contra una infección vírica. Un virus es un pequeño paquete de información genética, ADN o ARN, encerrado en una cápside, que consta de muchas copias de una o varias proteínas diferentes; la cápside, a su vez, puede aparecer envuelta en una membrana incrustada de proteínas. La infección comienza cuando un punto específico de la superficie del virus se une a un receptor específico de la superficie de la célula. El virus atraviesa la membrana celular y se apropia de la maquinaria biosintética de la célula para producir muchas copias de sí mismo; ello provoca la lisis de la célula y la liberación del virus, que habrá de iniciar otro ciclo infeccioso. El sistema inmune puede bloquear el proceso infeccioso si el huésped se ha expuesto anteriormente a ese mismo virus, o si se ha inmunizado contra él con una vacuna. En ambos casos, determinadas células del sistema inmune, denominadas linfocitos, se habrán estimulado con anterioridad para reconocer al virus. Cuando los receptores de uno de esos linfocitos encuentran al virus, se induce la respuesta inmunológica, que desencadena, entre otros procesos, la proliferación de células plasmáticas que secretan anticuerpos específicos para determinados puntos de la superficie del virus. Los anticuerpos se unen a esos sitios, recubren la superficie del virus y bloquean su

1. VACUNA SINTETICA diseñada con ayuda de un programa de representación gráfica por ordenador. Estas creaciones, de Arthur J. Olson, del Instituto de Investigaciones de la Clínica Scripps, muestran un método para identificar las regiones de una proteína que quedan sobre la superficie de un virus y son, por tanto, accesibles a los anticuerpos. Se representa (*1*) el esqueleto de la zona de la proteína próxima a la superficie de la envoltura externa del virus "achaparrado peludo del tomate", basándose en las coordenadas determinadas por Stephen C. Harrison y sus colegas, de la Universidad de Harvard. En amarillo (*2*), se destaca un péptido de la proteína, con las cadenas laterales de los aminoácidos que lo componen esquematizadas a nivel molecular. El péptido se ha ampliado (*3*) y se muestra una esfera que representa una molécula de agua. La esfera se hace girar alrededor del péptido para generar un mapa de la superficie accesible al agua (*4*); para ello se sigue un algoritmo desarrollado por Michael L. Connolly. En cada uno de los puntos de contacto más próximos al péptido se coloca una marca, para lo cual hay que considerar el radio de van der Waals de la esfera (zona de influencia efectiva) y el de cada átomo del péptido y el resto de la proteína. Se traza un mapa punteado similar de la superficie para mostrar las zonas del péptido que permanecen accesibles al agua después de asociarse tres copias de la proteína sobre la superficie del virus (*5*). Finalmente se reúnen cuatro de esas asociaciones sobre la superficie del virus (*6*).


1
2
3
4
5
6

2. PREVENCION DE LA INFECCION VIRICA por vacunación con el virus intacto o a través de un pequeño péptido artificial que reproduzca una zona de una proteína de superficie del virus. Al infectar la célula (*a*), el virus se une a un receptor de la misma; penetra en ella y se apodera de su maquinaria biosintética para obtener así muchas copias de sí mismo, lo que provoca la lisis celular. Un virus atenuado o inactivado puede inyectarse como vacuna (*b*). Cuando tal virus encuentra un receptor en un linfocito genéticamente programado para reconocer y producir anticuerpos contra una proteína de superficie del virus, el linfocito se estimula y forma un clon de células plasmáticas secretoras de anticuerpos. En cualquier exposición subsiguiente al virus, los anticuerpos se unen a la proteína y neutralizan el virus, impidiendo su unión a las células del huésped. Un péptido sintético, que mimetiza parte de la proteína de superficie, puede servir de vacuna (*c*), puesto que induce anticuerpos capaces de unirse a esa proteína.

unión a los receptores celulares. El virus queda neutralizado.

El sitio reconocido por un linfocito específico, y al que se unirá la zona de enlace al antígeno de un anticuerpo específico, se denomina determinante antigénico. Normalmente es una zona discreta de la superficie de una proteína vírica. ¿Podrían, pues, inducirse anticuerpos capaces de neutralizar a un virus, inyectando, como vacuna, ciertos péptidos que actuaran de determinantes antigénicos? La idea de que una pequeña parte de un agente infeccioso sirva de vacuna data de hace más de 45 años. Se remonta a los estudios realizados por Walther F. Goebel, del Instituto Rockefeller de Investigaciones Médicas, quien demostró que, inmunizando ratones con una cadena de azúcares obtenida de la superficie de la bacteria causante de la neumonía, podía protegerlos de infecciones posteriores. Las posibilidades de construir péptidos inmunogénicos a medida aumentaron extraordinariamente en la década de 1960, cuando R. B. Merrifield, de la Universidad Rockefeller, desarrolló un método para construir péptidos uniendo automáticamente aminoácidos, las subunidades de las proteínas.

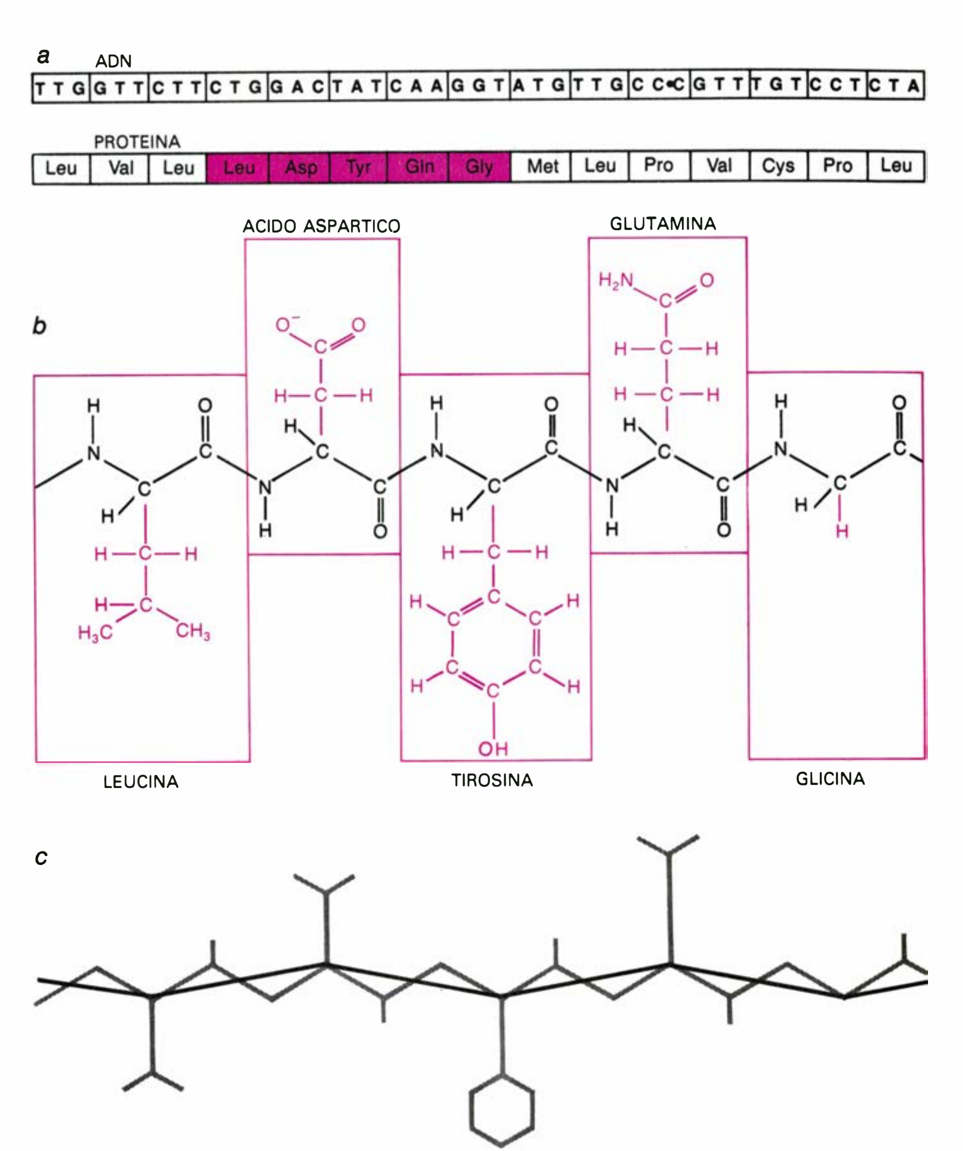
No siempre es fácil determinar la secuencia de aminoácidos de un determinado péptido incluido en una macromolécula proteica, pero se dio con una forma de evitar el escollo. La secuencia de aminoácidos de una proteína viene especificada por la secuencia de subunidades, llamadas nucleótidos, del ácido nucleico (ADN o ARN) que codifica tal proteína. El desarrollo, reciente, de poderosas técnicas para determinar las secuencias de nucleótidos del ADN ponía a nuestro alcance la posibilidad de determinar con rapidez la secuencia entera del ADN de un virus (o de una copia en ADN de un ARN vírico). En principio, podría identificarse la secuencia del ácido nucleico que codificara una proteína de la superficie y traducir esa región de acuerdo con el código genético, para obtener la secuencia de aminoácidos de la proteína. Luego, mediante la técnica de Merrifield, podrían sintetizarse péptidos cortos de varias regiones de la proteína, inyectarlos en animales de laboratorio y comprobar si inducían anticuerpos que se unieran a la proteína natural y, tal vez, neutralizaran el virus.

Antes de emprender el diseño de vacunas había que responder a una serie de difíciles cuestiones, relacionadas con la naturaleza del determinante antigénico y con la posibilidad de que

un péptido sintético corto lograra, o no, reproducirlo. Los anticuerpos se unen a las proteínas en función de su forma y carga eléctrica. El esqueleto de una proteína es una cadena de átomos de carbono y nitrógeno de la que se proyectan las cadenas laterales de la secuencia de aminoácidos, cada uno de los cuales tiene una forma y distribución de carga características. La secuencia de aminoácidos (que constituye la estructura primaria de una proteína) y sus consiguientes interacciones de corto alcance, determinan cómo gira la cadena para dar lugar a lo que se conoce como estructura secundaria: tal vez una hélice, o un pliegue pronunciado. Las interacciones entre las estructuras secundarias, a su vez, determinan el plegamiento de la cadena sobre sí misma para formar la compleja estructura terciaria, característica de cada proteína.

De ello se deduce que pueden existir dos tipos de determinantes antigénicos. Si el sitio de unión al antígeno de una molécula de anticuerpo reconoce y se enlaza a un grupo de aminoácidos contiguos, se trata de un determinante antigénico continuo. Un determinante discontinuo será el formado por grupos de aminoácidos que estén separados entre sí a lo largo de la cadena, pero que los plegamientos terciarios los aproxima.

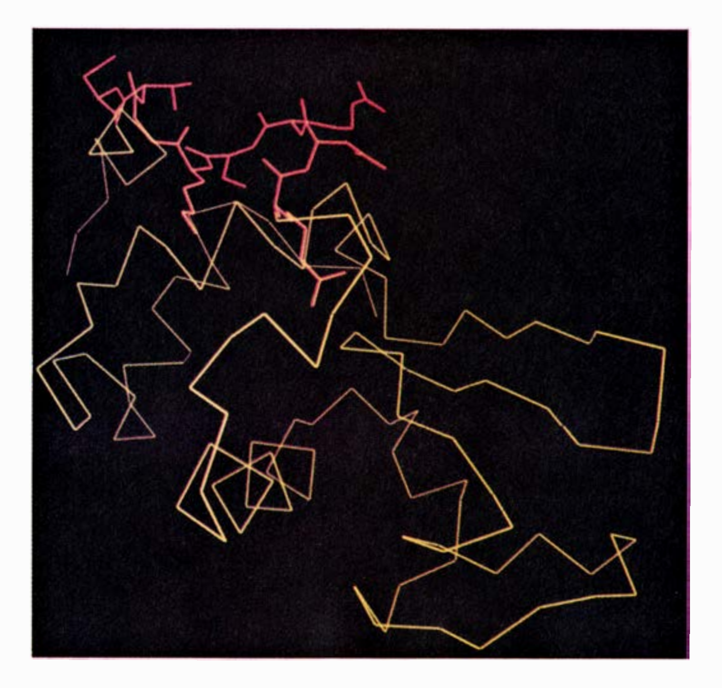
Hace dos o tres años, la mayoría de las doctrinas inmunológicas consideraban improbable que los péptidos sintéticos reprodujesen fielmente los determinantes antigénicos naturales de las proteínas víricas intactas. Parece, en efecto, que la respuesta inmune a una proteína vírica se dirige, en el curso de una infección natural, contra un pequeño número de determinantes antigénicos localizados en unas pocas zonas de la molécula proteica. Más concretamente, existen pruebas de que en la inmunogenicidad de una proteína (su capacidad para inducir la síntesis de anticuerpos) interviene sólo uno de cada 50 aminoácidos. Es más, la mayoría de estos determinantes antigénicos, poco abundantes, parecen localizarse de un modo discontinuo. Para lograr que un péptido sintético reproduzca un determinante de esa naturaleza habría que sintetizar un tramo largo de la proteína y, si bien esta síntesis es técnicamente posible, no es eficaz. Aunque se sintetizase un péptido largo, parece poco probable que, faltando el resto de la molécula, se plegase para dar la estructura terciaria correcta. De hecho, se pensó que las cadenas laterales individuales en un péptido corto no se orientarían espontáneamente por sí mismas, ni siquiera para reproducir la estructura secundaria más simple de un determinante antigénico continuo.

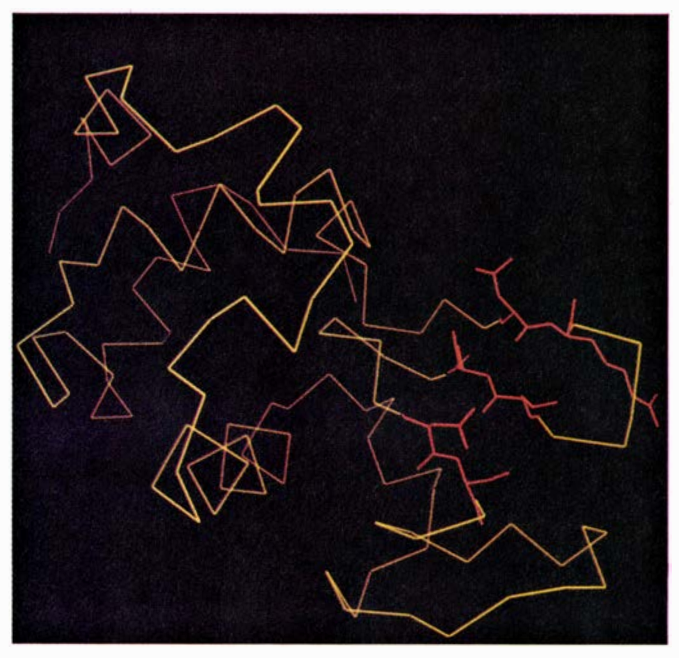


3. SECUENCIA DE NUCLEOTIDOS de un fragmento de ADN vírico que codifica o rige la secuencia de aminoácidos de una proteína del virus (*a*). Conocida la secuencia del ADN, se puede predecir la correspondiente secuencia de aminoácidos y seleccionar, para su síntesis, aquellos péptidos, o fragmentos de la cadena proteica, que con mayor probabilidad puedan inducir la síntesis de anticuerpos. Se detalla en *b* la zona del péptido que se indica en color en *a*. La cadena principal está constituida por átomos de carbono (C) y nitrógeno (N), con proyecciones de átomos de hidrógeno (H) y de oxígeno (O). Cada aminoácido se caracteriza por una cadena lateral (*color*), proyectada desde el llamado carbono alfa. En los diagramas de la proteína realizados por ordenador que ilustran este artículo, el esqueleto de carbonos alfa se muestra (*c*) en forma simplificada (*negro*) o a nivel atómico, si bien excluyendo los átomos de hidrógeno, (*gris*).

Para desafiar estas suposiciones, más bien pesimistas, decidimos hace dos años sintetizar, en fragmentos cortos, casi todos los tramos de una proteína vírica, y comprobar si los anticuerpos sintetizados contra los diferentes péptidos, podían o no reaccionar con la proteína completa. La molécula elegida fue la hemaglutinina de una estirpe del virus de la gripe. La hemaglutinina es una glicoproteína (es decir, una proteína que lleva adosadas cadenas de azúcares), que forma proyecciones radiales desde la envoltura esférica del virus. Las espículas de hemaglutinina poseen funciones importantes. Median la unión del virus a la célula huésped, son las responsables principales de la inducción de anticuerpos neutralizantes y, como su nombre indica, provocan el aglutinamiento de los eritrocitos. Constituyen, también, el factor principal de conducción de la síntesis de los anticuerpos que neutralizan el virus.

Dos logros recientes han hecho de la hemaglutinina del virus de la gripe objeto adecuado para nuestros estudios. Walter Fiers y sus colegas, de la Universidad belga de Gante, han determinado la secuencia de nucleótidos del gen que codifica esta proteína y, a partir de ella, han deducido la secuencia de aminoácidos. Ian A. Wilson y Don C.





4. LOS DETERMINANTES ANTIGENICOS son zonas de una proteína susceptibles de reconocimiento por parte de los anticuerpos, que se unen a ellos. Los dos tipos de determinantes que existen se ilustran en estas representaciones por ordenador; se muestra, en amarillo, parte de la cadena principal del enzima lisozima, con los determinantes antigénicos (cadenas laterales incluidas) en naranja. Un determinante de tipo continuo (*izquierda*) está constituido por una secuencia de aminoácidos contiguos situados a lo largo de un único segmento de la cadena. Un determinante de tipo discontinuo (*derecha*) es el compuesto por diversos segmentos separados, que el plegamiento terciario de la proteína sitúa en vecindad para formar un sitio antigénico único.

Wiley, de la Universidad de Harvard, y John J. Skehel, del Instituto Nacional de Investigaciones Médicas de Londres, han determinado la estructura de la proteína mediante cristalografía por rayos X, lo que nos ha permitido relacionar los distintos péptidos que hemos sintetizado con una región específica de la estructura secundaria y terciaria. La molécula de hemaglutinina consta de dos cadenas proteicas, denominadas *HA*1 y *HA*2: cada espícula es un trímero compuesto por tres moléculas de este tipo. *HA*1 parece ser el objetivo primario del sistema inmune, por lo que nos centramos en dicha cadena en primer lugar.

Nuestro grupo sintetizó 20 péptidos, el 75 por ciento de la longitud de la cadena aproximadamente; ofrecían una gran variedad de estructuras secundarias y terciarias. Cada péptido, acoplado a una proteína portadora, se inyectó en conejos; se valoró el suero de cada uno de los conejos (mediante un método que utliza un enzima unido a un anticuerpo) para determinar la presencia de anticuerpos que pudieran unirse al péptido inyectado, a la molécula de hemaglutinina y al virus intacto. Los 20 péptidos estudiados indujeron la formación de anticuerpos específicos; 15 de los anticuerpos antipéptido reaccionaron con la molécula completa de proteína y 17 de ellos se unían al virus de la gripe. Es decir, un péptido procedente de prácticamente cualquier región de la proteína vírica puede inducir un anticuerpo que reconozca a la proteína completa, tanto en forma purificada como sobre la superficie del virus. La inmunogenicidad natural de una proteína (la única bien conocida en la actualidad) es menor que el potencial inmunogénico de sus fragmentos. Parece que el único requisito es que el péptido sintetizado se halle sobre la superficie de la proteína plegada, de suerte que el anticuerpo pueda acceder a él.

Hay varias maneras de averiguar si determinado péptido se asienta en la superficie de una proteína; para decirlo con mayor exactitud, si un péptido (o una parte de él) es accesible a las moléculas de anticuerpo presentes en el suero de un animal. Podemos programar un ordenador, por ejemplo, para que dé una representación gráfica de las regiones accesibles al solvente [*véase la figura 1*]. En primer lugar se representa todo el péptido en estudio, con sus cadenas laterales en la conformación que adoptan en la molécula intacta. El ordenador hace rodar sobre el péptido una esfera, que representa una molécula de agua. La esfera penetra donde puede, en función del radio de van der Waals (aproximadamente, la zona de influencia en términos de carga eléctrica) de la molécula de agua y de los átomos de las cadenas laterales. La esfera deja una señal en cada punto de contacto con el péptido, dibujando un mapa de las regiones del péptido accesibles al solvente.

El esquema que he descrito es válido, pero exige el conocimiento de la estructura molecular del virus y de sus proteínas, datos desconocidos todavía para muchos virus. Afortunadamente, existe un procedimiento que, con sólo conocer la secuencia de aminoácidos, permite predecir qué fragmentos de una proteína es más probable que se localicen sobre la superficie. Los 20 aminoácidos se encuadran, en su mayoría, en dos grandes grupos: hidrófilos (con afinidad por el agua) e hidrófobos (que la repelen). Los aminoácidos hidrófobos suelen quedar confinados en el interior de la proteína plegada, mientras que los hidrófilos se localizan, por lo general, en la parte externa de la molécula. Los cuatro aminoácidos más hidrófilos son lisina, arginina, ácido aspártico y ácido glutámico. Se supone que las regiones de una proteína relativamente ricas en estos aminoácidos se hallarán expuestas a un entorno acuoso y, por tanto, se ubicarán en la superficie de la molécula.

Nuestros resultados con el virus de la gripe sugieren, como he indicado antes, que casi cualquier péptido de la superficie de la molécula inducirá anticuerpos capaces de reaccionar con la proteína entera. ¿Cómo es posible que un péptido aislado adopte una conformación similar a la que adoptaría cuando se pliega integrado en una molécula proteica completa? Quizá sea precisamente la ausencia de restricciones en su conformación lo que posibilite su geometría

correcta. Un péptido pequeño, inyectado en solución en el torrente circulatorio de un animal, se retuerce y pliega. Mientras cambia su conformación, prosigue, a buen seguro, interaccionando con los linfocitos; así se induce una serie de anticuerpos contra cada una de las conformaciones. De estas conformaciones, una debe coincidir muy estrechamente con la que el péptido adopta al formar parte de la molécula proteica entera. (Conviene destacar que, en pocos casos, los péptidos pequeños son mucho más eficaces que la proteína purificada a la hora de inducir anticuerpos contra la proteína vírica. Quizá la proteína aislada se pliegue en una configuración distinta de la que adquiere cuando se ordena junto con otras muchas copias idénticas sobre la superficie del virus.)

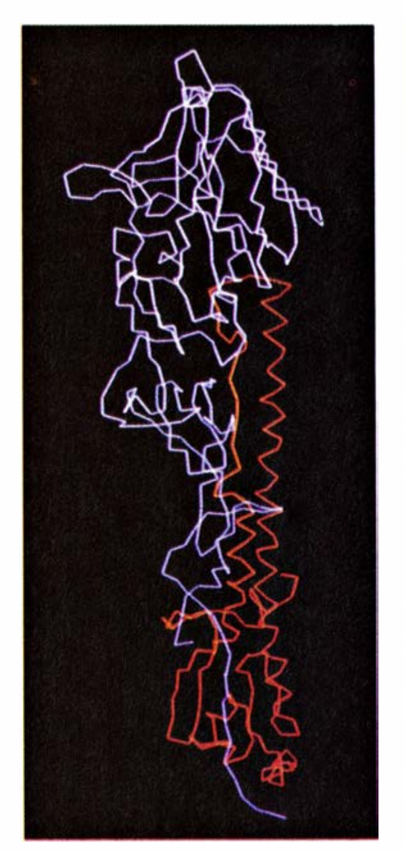
Si prácticamente cualquier región de la superficie de una proteína vírica induce la síntesis de anticuerpos contra esa proteína, queda abierto el camino para el diseño racional de vacunas, sin necesidad de inyectar partículas víricas y aceptar lo que el sistema inmune nos ofrezca. Este procedimiento se está aplicando ya en diversos laboratorios. La primera vacuna que nosotros desarrollamos lo fue contra el virus productor de la glosopeda, una epidemia histórica que afecta todavía al ganado, particularmente al vacuno, y que supone un serio problema en todos los continentes, con la excepción de Australia y América del Norte. Aproximadamente unos 3000 millones de dosis de vacuna de virus inactivado se administran cada año, en un esfuerzo por controlar la enfermedad. La envoltura del virus está formada por 240 moléculas proteicas, 60 copias de cada una de las cuatro proteínas víricas *VP*1, *VP*2, *VP*3 y *VP*4. Se ha intentado la síntesis de una vacuna basada en la proteína *VP*1, supuesto blanco natural más importante del anticuerpo neutralizante, pero se ha visto que la proteína entera no era muy inmunogénica. En colaboración con Fred Brown y David J. Rowlands, del Instituto de Investigación de Virus Animales de Inglaterra, emprendimos la síntesis de una subregión más inmunogénica de la proteína.

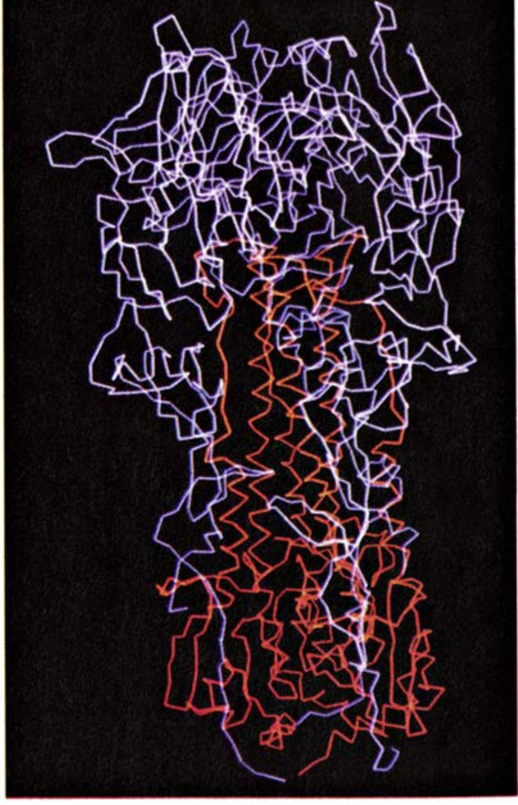
A partir de la secuencia de nucleótidos se acababa de determinar la secuencia de los 213 aminoácidos de la *VP*1. Aunque se desconocía la estructura de la proteína a nivel molecular, el análisis de la secuencia de aminoácidos bastó para centrar nuestra atención en la región comprendida entre los aminoácidos 141 y 160. Es esta una región particularmente rica en cadenas laterales hidrófilas de arginina, lisina y ácido aspártico, lo que sugería su posible localización en la superficie. La prolina se presenta, por lo común, donde la molécula proteica forma un plegamiento brusco (que puede inducir una respuesta inmunológica específica); en esta región había dos prolinas.

Y otro dato de mayor importancia: se sabía que, en esa región, la secuencia de aminoácidos variaba considerablemente de una estirpe de virus a otra. Un virus necesita eludir las defensas del sistema inmunológico del huésped durante el tiempo suficiente para invadir las células susceptibles y producir la progenie. Ello lo consigue cambiando continuamente los determinantes antigénicos de su cubierta más externa, o de la envoltura; así, los anticuerpos formados antes contra la estirpe parental no se unirán al virus modificado, ni lo neutralizarán. En términos menos teleológicos, lo que sucede es que la selección natural estimula la supervivencia de una estirpe en la que la mutación espontánea ha producido una variación importante desde el punto de vista inmunológico. Una región sometida a tal frecuencia de mutación probablemente sea la que acostumbra a inducir anticuerpos y, por tanto, un péptido de esa región constituye un buen candidato para una vacuna sintética.

Nuestro grupo sintetizó varios péptidos de la *VP*1 del virus de la glosopeda, incluyendo el péptido que se extiende desde la posición 141 hasta la 160; se acopló cada uno a una molécula portadora y se inyectaron en conejos. Tras determinar que los péptidos estimulaban la producción de anticuerpos específicos, realizamos un bioensayo para comprobar su capacidad neutralizadora del virus. Cuando el antisuero de conejo específico para el péptido 141-160 se inyectó en ratones, junto con el virus, lo neutralizó, según era de esperar; se había administrado a los ratones lo que se conoce como inmunidad pasiva mediante la inyección de anticuerpos antipéptido. Existen tres estirpes principales del virus de la glosopeda; nosotros habíamos trabajado con el tipo *O*. Cuando probamos el efecto de los anticuerpos sobre las otras estirpes, la neutralización fue mucho más débil, lo que apoyaba la creencia de que los cambios en esa región permitían al virus evitar los anticuerpos elaborados contra una estirpe diferente.

La prueba definitiva de cualquier



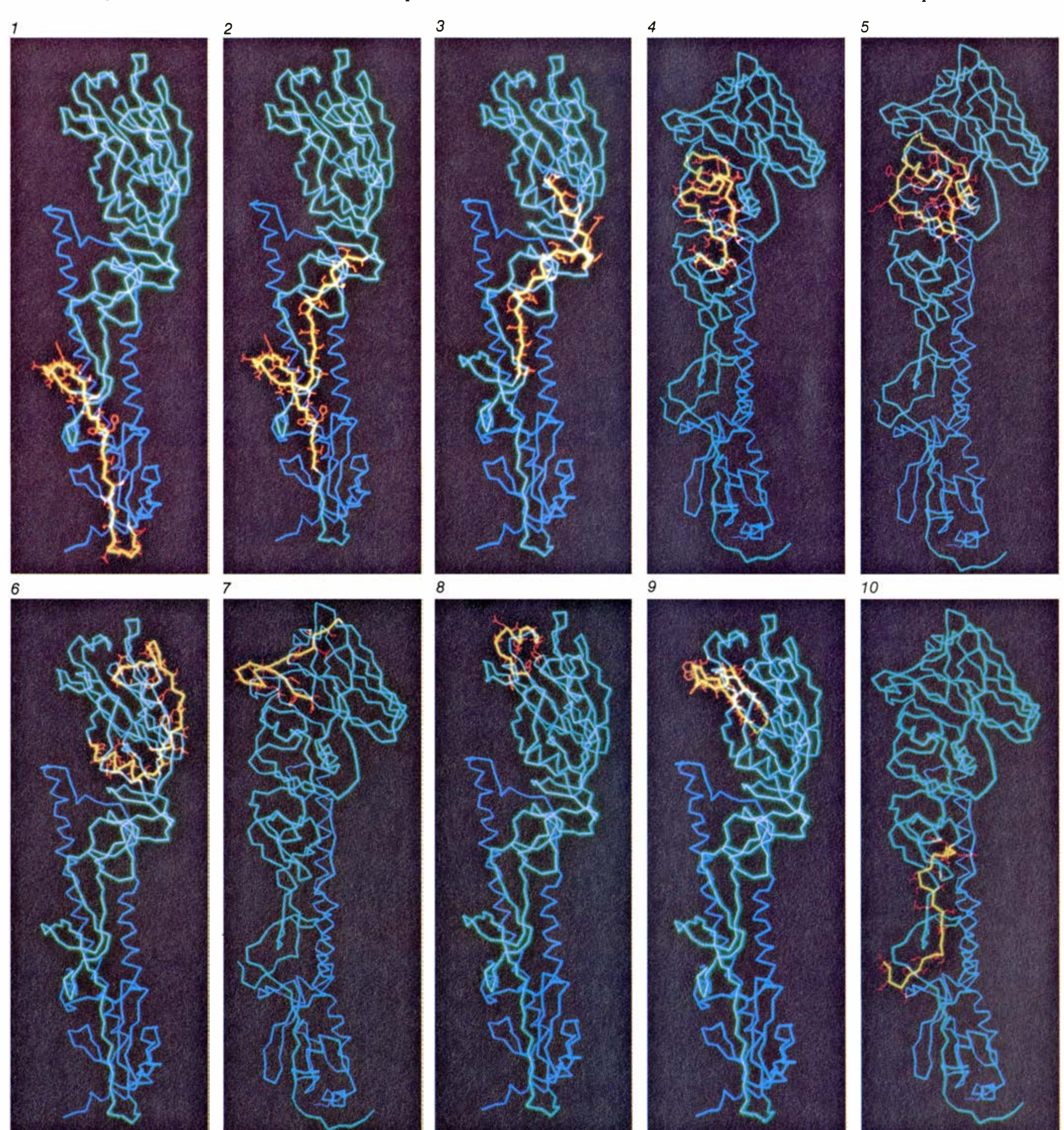


5. MOLECULA DE HEMAGLUTININA del virus de la gripe. Está constituida por dos cadenas polipeptídicas: *HA*1 (*morado*) y *HA*2 (*naranja*). La estructura tridimensional de la molécula (*izquierda*) se basa en los datos obtenidos por cristalografía de rayos X por Ian A. Wilson y Don C. Wiley, de la Universidad de Harvard, y John J. Skehel, del Instituto Nacional para la Investigación Médica de Londres; sólo se representa el esqueleto. En el trímero de hemaglutinina (*derecha*) se reúnen tres cadenas de *HA*1 y tres de *HA*2. Los trímeros forman las proyecciones situadas sobre la envoltura esférica de la partícula vírica.

vacuna es su capacidad de inducir inmunidad activa, además de inmunidad pasiva, es decir, de obligar a que el sistema inmune sintetice sus propios anticuerpos para protegerse de un futuro ataque. Inmunizamos cobayos con una inyección (200 microgramos) del péptido sintético; a los 35 días se les administró una fuerte dosis de virus: 10.000 veces la dosis requerida normalmente para infectar la mitad de un grupo de animales. Ninguno de los animales inmunizados enfermó, mientras que todos los animales no inmunizados murieron. Un ensayo similar de inmunidad activa se está llevando a cabo actualmente en Inglaterra con ganado vacuno. Ya hemos comprobado que el ganado inmunizado con el péptido produce anticuerpos a concentraciones que deberían protegerle contra una reinfección.

Obviamente, una vacuna de utilidad práctica deberá proteger contra las tres estirpes del virus. Hemos sintetizado péptidos de esa misma región (141 a 160) de la *VP*1 de los virus tipo *A* y tipo *C*, que tienen una secuencia de aminoácidos ligeramente distinta. Hasta ahora los anticuerpos contra esos péptidos se han ensayado en ratones y neutralizan los virus de los tipos *A* y *C*, respectivamente. No está muy lejos la fabricación de la vacuna sintética contra los tres tipos del virus de la glosopeda.

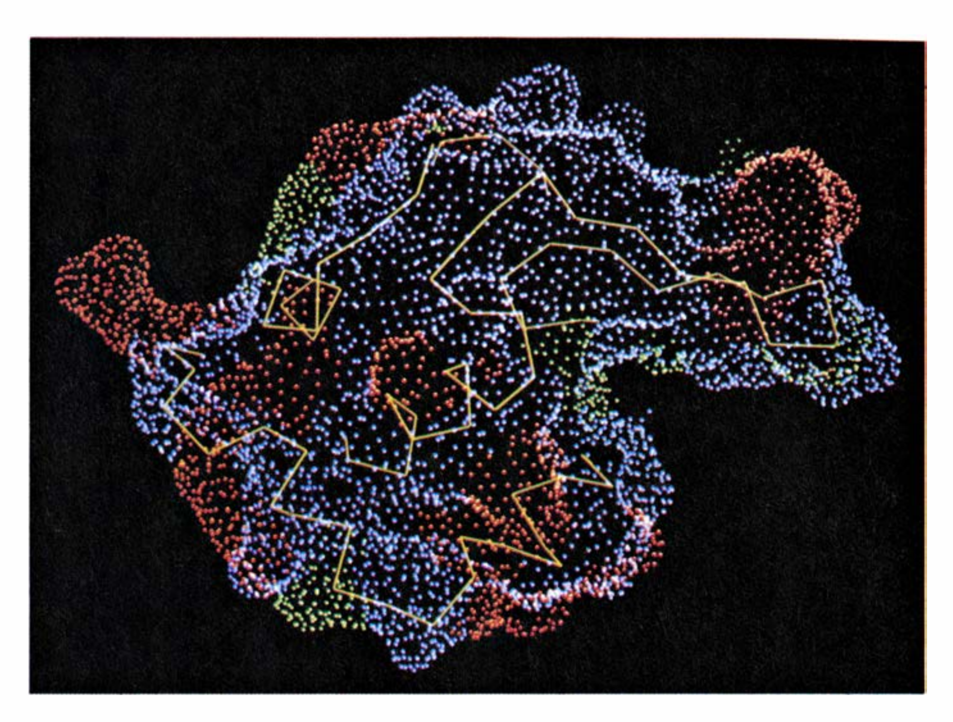
Dado que, en principio, se ha resuelto la cadena técnica que va desde la



6. SERIE DE PEPTIDOS sintetizada para ensayar en ellos la capacidad de inducir anticuerpos que reaccionen con los péptidos mismos, con la molécula entera de hemaglutinina y con la partícula vírica. La serie totaliza el 75 por ciento de la cadena de *HA*1. Diez de los veinte péptidos ensayados se muestran, en amarillo, en esta representación; en naranja aparecen sus cadenas laterales; el resto de la cadena *HA*1 se ofrece en azul claro y, la cadena *HA*2, en azul oscuro. En cada caso se ha girado el monómero de hemaglutinina para mostrar mejor el péptido correspondiente. La mayoría de los péptidos indujeron, en conejos, anticuerpos que reaccionaron satisfactoriamente con el virus, lo cual apoya la pretensión de utilizar péptidos sintéticos como vacunas.

secuencia del ácido nucleico hasta el péptido sintético, el siguiente paso a dar será el de investigar con más detalle ciertos aspectos del diseño de las vacunas. Las investigaciones actuales se dirigen hacia la búsqueda de formas nuevas y más refinadas de definir las regiones de las proteínas víricas más vulnerables al ataque inmunológico, así como la manera de inducir una inmunidad duradera.

En las proteínas víricas se distinguen dos tipos principales de regiones. En una de ellas se permite la variación; es más, ésta es esencial si el virus quiere confundir al sistema inmunológico del huésped. Una vacuna que dé lugar a la formación de anticuerpos contra esa región (*HA*1 del virus de la gripe o *VP*1 del virus de la glosopeda) probablemente sea inmunogénica, pero además debe diseñarse para que elimine todas las estirpes variantes. En las proteínas víricas hay otras regiones que tienen una función biológica esencial y constante, cuya estructura no puede variar. La secuencia de aminoácidos de dichas regiones se conserva estrictamente en todas las estirpes del virus. Si una región esencial y permanente es accesible a los anticuerpos, la vacuna que induzca la producción de anticuerpos contra ella será universal.



7. REGIONES ACCESIBLES AL SOLVENTE en la molécula de lisozima. Se cartografiaron siguiendo el método que se muestra en la figura 1, pero con los puntos coloreados codificados de acuerdo con el aminoácido responsable en cada zona de la superficie: naranja para los aminoácidos básicos lisina y arginina, verde para los aminoácidos ácidos aspártico y glutámico y azul para todos los demás. Los aminoácidos básicos y ácidos son los más hidrófilos (con avidez por el agua). Como muestra su disposición, tienden a instalarse en la superficie. Si no se conoce con detalle la estructura de una proteína, cabe suponer que los péptidos ricos en aminoácidos hidrófilos serán los de mayor actividad inmunogénica.

El virus de la gripe es conocido por su capacidad para eludir el sistema inmunológico cambiando zonas de la hemaglutinina de su superficie. Los cambios se localizan en la cadena *HA*1; fueron los péptidos de esa cadena los primeros donde ensayamos la inmunogenicidad en el experimento descrito anteriormente. La hemaglutinina tiene, al menos, otras dos funciones que no varían. Forma una oquedad que se une al ácido siálico, un azúcar que se encuentra en la superficie de las células y que, por tanto, inicia el ciclo infectivo. Asimismo favorece la fusión de las membranas de las células infectadas, aumentando la infectividad vírica. Diversos estudios, realizados por Purnell W. Choppin y sus colegas, de la Universidad Rockefeller, y por Michael Waterfield, de la Fundación Imperial para la Investigación del Cáncer, de Inglaterra, sugieren que una región persistente situada al principio de la cadena *HA*2 es responsable de la fusión, al menos en parte.

Normalmente el sistema inmune no "ve" esta región, y no se producen anticuerpos contra ella en el curso de la infección natural. Si se dieran naturalmente estos anticuerpos (es decir, si la región no sólo fuera altamente persistente y biológicamente esencial, sino también muy inmunogénica), todos los seres humanos, y los demás animales hospedantes, serían, desde hace tiempo, inmunes al virus de la gripe: no existiría la gripe. Decidimos estudiar si un péptido sintético de esa región podía inducir anticuerpos que neutralizasen más de una estirpe del virus.

Stephen Alexander, Hannah Alexander y Nicola Green sintetizaron varios péptidos del principio de la molécula de *HA*2; eligieron la secuencia de aminoácidos del virus de la gripe tipo *A*, subtipo H_3. Antisueros de conejo contra estos péptidos neutralizaron el virus *A*, H_3 al ensayarlos en un cultivo celular, demostrando que, aunque esta región no suele ser inmunogénica, sí es accesible a los anticuerpos antipéptidos. Además, los mismos antisueros neutralizaron virus de un subtipo distinto, e incluso de un tipo diferente, como es el virus *B* de la gripe. En cambio, los anticuerpos inducidos contra el subtipo H_3 inalterado no neutralizaron otros subtipos. Los animales inmunizados con un subtipo de la gripe tampoco están protegidos contra la infección por otro subtipo. Queda aún mucho por averiguar de los anticuerpos anti-*HA*2, pero estos primeros resultados sugieren hasta qué punto hay que confiar en las vacunas destinadas a atacar una región persistente que desempeñe una función biológica importante.

Hemos visto, pues, que una forma de diseñar una vacuna es elegir, para su síntesis, una región de una proteína que sirva teóricamente como determinante antigénico. Hay otra manera de proceder. Se puede tratar de imitar determinantes antigénicos de los que ya se hubiese demostrado su importancia por estudios seroepidemiológicos. (En éstos, los virus íntimamente relacionados se clasifican de acuerdo con los anticuerpos que inducen.) Estos estudios muestran la existencia de diferentes tipos y subtipos de un virus, como los que he mencionado en el análisis de la glosopeda y la gripe.

Los tipos se definen por estudios a largo plazo, en los cuales los virus responsables de una determinada enfermedad se recogen en diferentes áreas y a intervalos temporales distintos. Se analizan los anticuerpos procedentes de diversos pacientes frente al panel de virus; cada anticuerpo reacciona con una estirpe vírica diferente: reconocen determinantes antigénicos diferentes. Los serólogos pueden describir un tipo de virus según su panel de determinantes antigénicos, que obtienen poniendo en parangón un elevado número de anticuerpos con un panel de virus. Pueden así saber cuál de esos determinantes ha funcionado en un determinado individuo o en el curso de una epidemia. Sin embargo, hasta ahora, los determinantes eran marcadores abs-

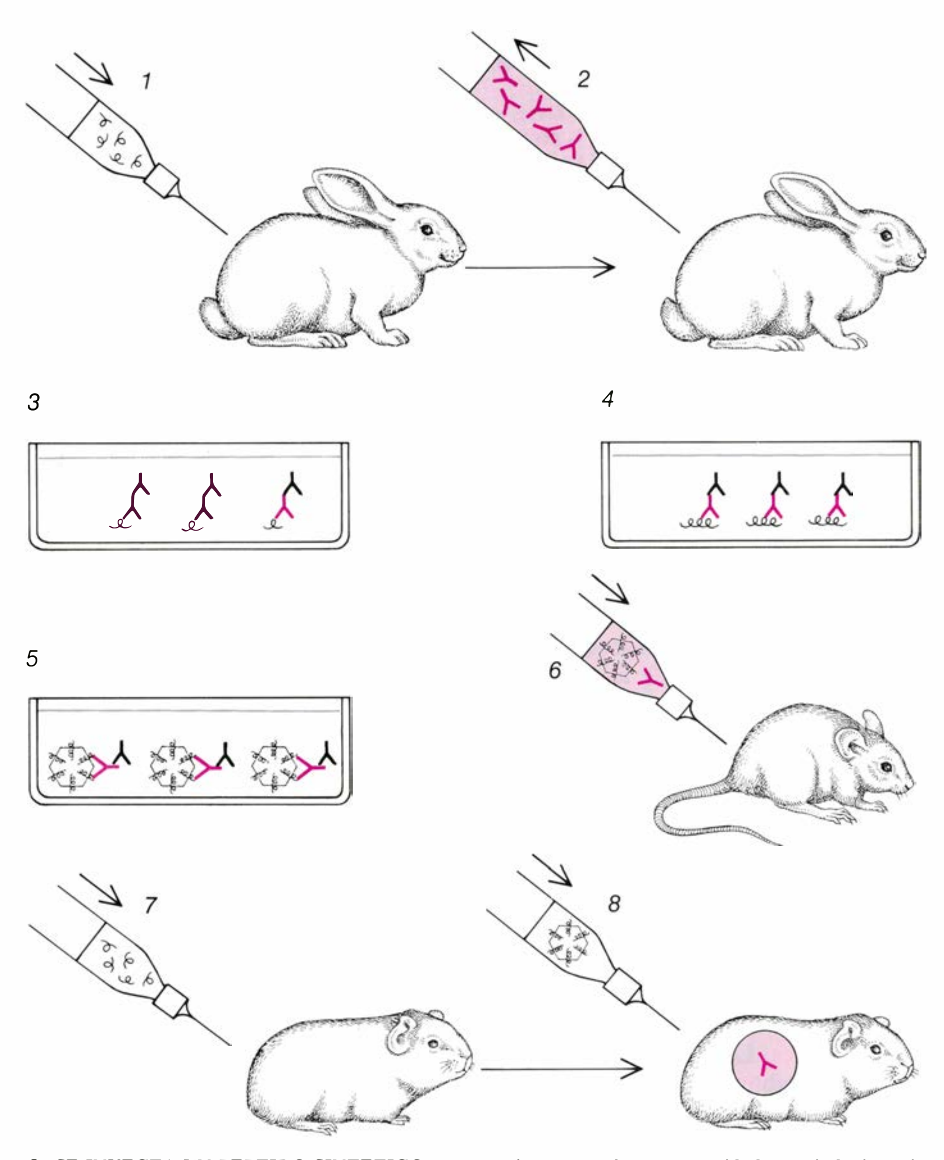
tractos, designados por letras o números; no se conocían a nivel molecular. No estaba claro que el cúmulo de datos seroepidemiológicos recogido facilitara el diseño de vacunas. Cabía que los péptidos sintéticos no reprodujesen las estructuras responsables de los marcadores serológicos, en el caso de que esas estructuras fueran discontinuas en la naturaleza. Decidimos intentar el diseño de una vacuna contra una enfermedad del hígado, la hepatitis *B*. Y sintetizamos péptidos que fueran molecularmente equivalentes a los determinantes serológicos abstractos de importancia reconocida.

El componente vírico que parece actuar de blanco principal de los anticuerpos neutralizantes, el antígeno de superficie del virus de la hepatitis *B*, es una proteína de la envoltura de la partícula vírica. Estudios serológicos demuestran que varias estirpes del virus de la hepatitis *B* poseen un determinante común, designado *a*. Cada estirpe tiene además otros dos determinantes: *d* o *y*, *w* o *r*; ello significa que pueden existir al menos cuatro tipos de virus: *adw, ayw, adr* y *ayr*. Trabajando en colaboración con John L. Gerin, de la Facultad de Medicina de la Universidad de Georgetown, y Robert H. Purcell, del Instituto Nacional de Alergia y Enfermedades Infecciosas de los Estados Unidos, tratamos de sintetizar péptidos que indujeran anticuerpos específicos para los determinantes *d* y *e*. Sin embargo, el antígeno de superficie de la hepatitis *B* es una cadena de 226 aminoácidos; necesitábamos, pues, una orientación para saber por dónde empezar. El análisis de la secuencia de aminoácidos del antígeno de superficie de tres estirpes del virus, realizado por varios investigadores, sólo había mostrado la existencia de diferencias importantes, entre las estirpes de virus, en la región de la proteína comprendida entre los aminoácidos 110 y 140. Cualquier diferencia entre determinantes debe originarse por diferencias en las secuencias de aminoácidos; estaba claro que ésa era la región donde debíamos buscar nuestros péptidos.

Siguiendo una de las tres secuencias de que se disponía, sintetizamos varios péptidos que abarcaban esa región y produjimos anticuerpos contra cada uno de ellos. Un péptido de 13 aminoácidos (de la posición 125 a la 137) inducía anticuerpos que sólo reaccionaban con virus que poseían el determinante *y*. Significaba ello que el determinante *y*, y su contrapartida, el determinante *d*, debían localizarse en el péptido de 13 aminoácidos. Lograda la síntesis de *y*, abordamos la del determinante *d*, utilizando alguna otra de las secuencias de esa región. La secuencia elegida difería de la primera sólo en tres posiciones: dos treoninas y una tirosina aparecían sustituidas por prolina, asparagina y fenilalanina, respectivamente. Los anticuerpos obtenidos contra ese nuevo péptido sólo reaccionaban con virus que poseían determinantes *d*.

Quizás el cambio de *y* a *d* se produzca por sustitución de menos de tres aminoácidos en la cadena polipeptídica. Sintetizando péptidos en los cuales sólo se cambia una de las tres posiciones cada vez, pronto podríamos saber si se requería una, dos o las tres sustituciones. Hasta ahora estos experimentos nos han enseñado dos cosas: un cambio sutil en la composición de aminoácidos de una proteína pueden definir una estirpe distinta de virus; y la segunda, los péptidos sintéticos pueden reproducir las diferencias puestas de manifiesto por los estudios serológicos.

Para que las vacunas sintéticas constituyan una alternativa de las convencionales debe aumentarse la magnitud y duración de la inmunidad que confieren. Una ventaja de los virus atenuados, como he mencionado, es que conservan la capacidad de proliferar y, por ello, no cesan en la presentación de antígenos al sistema inmune, aumentando así el nivel de anticuerpos. Un péptido sintético, al igual que un virus inactivado, no se replica y, para aumentar el nivel y la duración de la inmunidad, se



8. SE INYECTA UN PEPTIDO SINTETICO a un conejo para probar su capacidad para inducir anticuerpos (*1*). En suero extraído del conejo (*2*) puede valorarse la presencia de anticuerpos específicos para el péptido inyectado (*3*), para toda la proteína vírica (*4*) o para el virus (*5*). La unión de los anticuerpos antipéptido (*color*) se detecta por la actividad de un enzima marcador unido a un segundo anticuerpo (*negro*), que se une al primero. La capacidad del anticuerpo antipéptido de neutralizar el virus (para conferir inmunidad pasiva), se evalúa inyectando virus y antisuero a ratones (*6*); si el ratón sobrevive, el virus ha sido neutralizado. La capacidad del péptido sintético para servir como vacuna induciendo inmunidad activa puede probarse inyectando el péptido en cobayos (*7*) y más tarde el virus (*8*). Si el animal sobrevive (mientras que los animales no inmunizados mueren) la eficacia de la vacuna queda demostrada.

hace necesaria la adición de adyuvantes a la vacuna.

Entre los adyuvantes más eficaces se cuenta el de Freund, constituido por una emulsión de aceite mineral y agua, mezclada con bacterias muertas (micobacterias, una clase que incluye el germen de la tuberculosis). Aunque se desconoce el mecanismo de acción del adyuvante, se le supone una acción doble: por un lado, la emulsión retiene el antígeno, formando un depósito del cual se libera muy lentamente, simulando un estímulo antigénico mantenido, como en el caso de virus atenuado; por otro, las bacterias muertas atraen hacia el depósito diversas células que intervienen de forma notable en la respuesta inmune. La atracción de tales células sirve de irritante, lo que provoca la inflamación del lugar de la inyección e incluso produce abscesos. Los adyuvantes potentes, como el de Freund, no son aptos para los seres humanos; en la vacunación, se les suele sustituir por virus inactivados, por alúmina o hidróxido de aluminio. El antígeno se adsorbe en las partículas de hidróxido de aluminio y se libera lentamente después de la inyección. En algunas casos podrá utilizarse alúmina con vacunas sintéticas, si bien la mayoría de ellas necesitan algo más eficaz.

François Audibert y Louis Chedid, del Instituto Pasteur, y Ruth Arnon y Michael Sela, del Instituto Weizmann de Ciencias, acaban de conseguir anticuerpos que confieren inmunidad pasiva frente a la toxina diftérica. Han sintetizado tres péptidos correspondientes a la molécula de la toxina. Antes de su inyección hay que acoplar los péptidos sintéticos a moléculas portadoras (en todos nuestros experimentos la molécula portadora fue la hemocianina de las lapas con bocallave, un pigmento respiratorio de los moluscos que se utiliza con frecuencia para este propósito). El grupo Pasteur-Weizmann diseñó una combinación portadora-adyuvante. Acoplaron los péptidos de la toxina diftérica que habían producido a un portador al que previamente habían unido un derivado sencillo de la membrana celular de micobacterias; aparentemente, el derivado de la membrana sirve de adyuvante, aumentando significativamente la inmunogenicidad de los péptidos.

Como he mencionado al principio de este artículo, los anticuerpos específicos para los péptidos deberían tener aplicaciones importantes como instrumentos de investigación. Una de estas aplicaciones surge de los avances en la tecnología de ácidos nucleicos. Debido a la precisión de las técnicas de recombinación de ADN y a la rapidez con que se puede determinar la secuencia de un ácido nucleico, se da el caso frecuente de conocer antes la secuencia entera de un gen que la secuencia de la proteína que codifica: se tiene un gen en busca de la proteína. A partir de las secuencias del ADN o ARN podemos predecir la secuencia de aminoácidos y seleccionar una pequeña región de la misma para sintetizarla. Los anticuerpos inducidos por el péptido sirven de sonda para buscar la proteína. Así, por ejemplo, se han identificado y localizado en determinadas zonas de la célula, por este método, las proteínas codificadas por algunos oncogenes, genes de algunos virus inductores de tumores.

La característica más notable de los anticuerpos antipéptidos es, obviamente, la predeterminación de su especificidad. Son anticuerpos dirigidos por el investigador contra una región concreta y conocida de la proteína y no contra la proteína entera. Una vez producidos por el sistema inmunológico de los animales de experimentación, los anticuerpos pueden purificarse por un solo paso en columnas de afinidad. Se emplean luego para investigar la estructura pormenorizada de una proteína y correlacionar estructura y función. Por ejemplo, anticuerpos específicos para una región determinada de una proteína se unirán a esa zona e inhibirán su función; de este modo, los ensayos con una colección de péptidos nos permitirán definir el sitio activo de una proteína o los diversos sitios que son activos en las distintas funciones. Análogamente, muchas hormonas peptídicas se sintetizan como una cadena proteica larga que, posteriormente, cortarán los enzimas en pequeños péptidos, cada uno de éstos con una actividad hormonal específica. Los anticuerpos específicos para los péptidos permiten separar los fragmentos y correlacionar cada uno con su actividad.

Finalmente, los anticuerpos de especificidad predeterminada pueden aclarar qué porción de un gen se está expresando para sintetizar una proteína en un determinado momento. Los genes de los mamíferos aparecen fragmentados en una serie de regiones codificadoras denominadas exones. En el caso de ciertas proteínas variables, se expresan diferentes exones a diferentes tiempos. En el curso de una respuesta inmune se sintetizan distintas moléculas de anticuerpos. Podemos ya sintetizar el péptido codificado por cada exón, preparar anticuerpos específicos para él y averiguar qué exón se expresa en cada etapa de la respuesta inmune.

Ciencia y sociedad

Genes de compatibilidad

Los antígenos de histocompatibilidad, moléculas incrustadas en la membrana superficial de las células, imprimen a los tejidos de los animales su identidad química. Al trasplantar un colgajo de piel o un órgano, el sistema inmune del receptor reconoce como extraños esos antígenos, y procede a la destrucción del tejido injertado. El rechazo se manifiesta más tajante cuando donante y receptor poseen distintas versiones de ciertos genes: los del complejo mayor de histocompatibilidad, que en el hombre se sitúan en un breve segmento del cromosoma 6, denominado región *HLA*. En el ratón, unos genes análogos ocupan la región *H*-2 del cromosoma 17. La persistencia de los genes de histocompatibilidad a lo largo de la evolución no se debe a una voluntad expresa de que fracasen los trasplantes. Se sabe que gobiernan una amplia gama de funciones biológicas. Parece que determinan incluso los olores corporales por los que se reconocen, unos a otros, los animales.

En 1974, personal adscrito a la cría de ratones con fines experimentales del Memorial Sloan-Kettering Cancer Center observaron que el comportamiento social de éstos guardaba relación con su tipo *H*-2. Mantenían en varias jaulas ratones machos endogámicos con dos hembras. Una pertenecía a la misma cepa endogámica que el macho, es decir, era singénica (genéticamente idéntica); la otra era congénica para *H*-2, idéntica al macho en toda su información genética salvo en la región del cromosoma 17 que porta los genes *H*-2. El macho y la hembra que diferían en los genes *H*-2 tendían a aparearse y criar, marginando a la hembra singénica. Dábase la casualidad de que el entonces presidente del Memorial Sloan-Kettering, Lewis Thomas, había escrito poco antes sobre la posibilidad de que los genes de histocompatibilidad confirieran a los animales un olor característico. La estimulante coincidencia de las observaciones del trío de cría y la hipótesis de Thomas animó a los investigadores del centro, quienes formaron un grupo de trabajo con otros del Monell Chemical Senses Center, de Filadelfia, al objeto de estudiar si tales preferencias de apareamiento tenían algo que ver con los *H*-2 y el reconocimiento por el olor.

Empezaron por un ensayo controlado de la observación del trío de cría. Se enjauló un macho con una hembra singénica y otra congénica para *H*-2. Miles de ensayos más confirmaron la existencia de una desviación global a favor del cruce entre ratones con genes *H*-2 distintos. Tales cruzamientos daban descendencia heterozigótica (es decir, la progenie portaba juegos distintos de genes *H*-2 en las dos copias del cromosoma 17). Probablemente se trata de un carácter que redunda provechosamente en su eficacia biológica.

El paso siguiente a acometer consistía en establecer la forma de reconocimiento del *H*-2, que se suponía olfativa, pero que bien podía haberse fundado en el sabor, o incluso en el oído, pues se sabe que los ratones se comunican por señales ultrasónicas. Se estudió su capacidad para distinguir entre dos tipos *H*-2 en un laberinto en forma de *Y*. Se colocaba el macho en el tronco de la *Y*; los brazos de la *Y* se perfumaban con aire procedente de sendas cámaras donde se alojaba, en cada una, otro macho, o su orina. Los dos machos que proporcionaban el olor procedían de cepas congénicas para *H*-2. Se recompensaba con una gota de agua a los sedientos ratones sometidos a ensayo cuando se introducían en el brazo perfumado con el olor de un ratón de un determinado tipo *H*-2, o de su orina. El experimento se controló con gran precisión para evitar señales extrañas y la ayuda involuntaria de quienes manejaban el laberinto. Todos los animales ensayados, sin distinción de su linaje genético, aprendieron a distinguir entre los tipos *H*-2 que se les presentaron.

Los experimentos efectuados en el laberinto dieron pie a numerosas conclusiones, de las que informaron, en una reunión de la Sociedad Internacional de Biología y Medicina del Desarrollo Oncogénico, Edward A. Boyse, Gary K. Beauchamp, Kunio Yamazaki, Judith Bard y Thomas. Ha quedado confirmado el reconocimiento sensorial de los tipos *H*-2. Ante el éxito obtenido en el entrenamiento para distinguir sólo orina, se admite que el olfato constituye un modo importante de reconocimiento (aunque no necesariamente el único). El reconocimiento no depende del sexo. Pese a que los ratones que eligen pareja optan, en general, por los animales de un tipo *H*-2 distinto, no se observó en el laberinto, con ratones no entrenados, desviación hacia un tipo similar ni hacia otro distinto.

Prosiguen los estudios en el Monell y en el Memorial Sloan-Kettering, y van abriéndose nuevos interrogantes. ¿Hay más genes implicados en la definición del olor corporal? Sabemos ya que la variación en la región *Qa: T1a*, genéticamente relacionada y adyacente a *H*-2, marca de forma diferencial la orina de los ratones. ¿Qué efectos puede ejercer la región *H*-2 sobre la conducta reproductora, aparte de la preferencia en el cruce? Hay pruebas de que afecta al fenómeno conocido por bloqueo del embarazo. La implantación del óvulo fecundado puede fracasar cuando se expone una hembra a un macho extraño poco después de la cubrición; se ha probado que el bloqueo del embarazo es más probable si el macho extraño es congénico para *H*-2 con el macho semental. ¿Existe alguna relación entre el reconocimiento olfativo y el reconocimiento inmunológico? Parece que los genes del complejo mayor de histocompatibilidad están involucrados en ambos sistemas; el olfato, como la respuesta inmune, depende de receptores superficiales. En todo caso, se desconoce la naturaleza de los receptores olfativos y de las señales que envían al cerebro.

Bajo el volcán

Se afirma desde hace tiempo que las grandes erupciones volcánicas producen efectos sobre el clima mundial que duran varios años. La inyección de polvo y ceniza a la atmósfera reduciría la cantidad de radiación solar que alcanza la superficie terrestre, de lo que resultaría un enfriamiento global del planeta. No se ha determinado todavía el mecanismo exacto del fenómeno. Debido en parte a esta falta de conocimiento, la explosión del volcán mexicano El Chichón, ocurrida el año pasado, llamó poderosamente la atención de geólogos y meteorólogos. Michael R. Rampino, del Instituto Goddard de Estudios Espaciales de la NASA, y Stephen Self, de la Universidad estatal de Arizona, sostienen que el efecto de El Chichón sobre el clima mundial podría equiparse al de otras erupciones mucho mayores, que rebajaron las temperaturas anuales medias en el hemisferio Norte unos 0,4 grados Celsius.

Antes de su erupción, El Chinchón alcanzaba los 1400 metros de altura; se encuentra en el estado de Chiapas, el más meridional de México. En Chiapas se da la intersección de tres grandes placas tectónicas: norteamericana, caribe y placa de Cocos, que subyace al océano Pacífico, a la altura de Centroamérica. La subducción de la placa de Cocos bajo México generaría actividad volcánica en un amplio cinturón del sur

de México. El Chichón se encuentra en el extremo más sureño de esa zona. Mantenido en calma a lo largo de toda la historia, explotó, con gran violencia, el 28 de marzo, repitiendo la erupción el 3 y el 7 de abril.

Inmediatamente después de cada una de esas tres principales erupciones aparecía, en fotografías de México tomadas desde satélite, una gran nube de polvo, ceniza y gas. Las tres nubes, que alcanzaban los 16 kilómetros de altura, no tardaron en fundirse en una sola masa. La parte inferior de esa nube única se desplazó al nordeste, hacia el golfo de México; la deposición de partículas de polvo y la acción erosiva de las nubes de agua fueron debilitando su grosor. La parte superior de la nube, que llegaba hasta la estratosfera, se encaminaba al oeste, sobre el Pacífico.

A mediados de abril, la masa estratosférica se detectó sobre Hawai, mediante determinaciones de la cantidad de radiación que llega a la tierra. En mayo, la nube se observó sobre Ahmadabad, India, y se extendía desde una altitud de 17 kilómetros hasta los 30 kilómetros, abarcando casi la totalidad de la estratosfera. Se supone que la nube estratosférica rodeó la tierra en una estrecha faja y se difundió luego a latitudes superiores e inferiores hasta cubrir el globo entero.

¿En qué medida afectará la nube de El Chichón a la climatología? Rampino y Self proponen, en *Quaternary Research*, correlacionar la cantidad y tipo de materiales expulsados en tres grandes erupciones volcánicas con los subsiguientes cambios de la temperatura anual media registrados en el hemisferio Norte. El Tambora, en la isla indonésica de Sumbawa, entró en actividad en 1815 y provocó un enfriamiento de 0,8 grados Celsius; el Krakatoa, en Sunda Straits, entre Java y Sumatra, lo hizo en 1883 y redujo la temperatura media unos 0,4 grados. El monte Agung, de Bali, despertó en 1963 y enfrió el planeta otros 0,4 grados.

Para explicar el efecto de los volcanes sobre el clima suele considerarse el volumen total de materiales expulsados durante la erupción. Sin embargo, aunque el efecto de las erupciones del Tambora, Krakatoa y monte Agung sobre el clima fue similar, el volumen de rocas expulsadas resultó muy distinto: unos 150 kilómetros cúbicos en el caso del Tambora, 20 kilómetros cúbicos en el Krakatoa y un kilómetro cúbico en el monte Agung. Para explicar el desproporcionado efecto de la erupción de este último Rampino y Self se basan en la composición andesítica del volcán. (La andesita es una roca volcánica cuya composición queda a medio camino entre los tipos basáltico y ricos en silicio.) La roca andesítica suele contener más azufre que las rocas ricas en silicio del Tambora y Krakatoa. El análisis de muestras de roca prueba que el contenido en azufre del monte Agung es de unas 800 partes por millón, mientras el Tambora tiene 380 partes por millón y el Krakatoa 150.

En una erupción, el azufre se convierte en dióxido de azufre gaseoso, que reacciona con el vapor de la estratosfera y rinde un fino aerosol de ácido sulfúrico. Rampino y Self señalan que parece haber un límite superior en la cantidad de aerosol ácido que puede formarse en la estratosfera; de ahí que un gran incremento de la cantidad de materia expulsada no provoque un aumento correspondiente de la cantidad de ácido. Es más, por encima de los 30 kilómetros de altitud las gotitas se evaporan rápidamente, por lo que una erupción de potencia excepcional, que enviara material a elevadas altitudes de la atmósfera, generaría menos aerosol que otra menos violenta.

Han estimado la cantidad de aerosol con azufre inyectado a la atmósfera por los volcanes midiendo la cantidad de sulfato depositado en los estratos anuales del casquete polar. Las mediciones muestran que la producción de aerosol a partir del Tambora, Krakatoa y monte Agung se ajustaban a la proporción 7,5:3:1, muy inferior a la proporción del volumen total de materiales expulsados. Además, dada la imprecisión de los cálculos, la proporción de sulfato se acerca a la relación de reducciones de temperaturas observadas. Concluyen que el aerosol ácido condiciona el bloqueo de la radiación solar.

La erupción de El Chichón fue pequeña, al lado de esos tres ejemplos históricos. Expulsó unos 0,5 kilómetros cúbicos de roca. Sin embargo, como el monte Agung, El Chichón es un volcán andesítico, y además, según Rampino, "excepcionalmente rico en azufre", más rico aún de lo que pudiera indicar el tipo de roca. "Quizás entró azufre procedente de sulfatos contenidos en rocas que subyacen al volcán", añade. Las muestras tomadas cerca del volcán sugieren la existencia de un estrato cargado de sulfato cálcico.

Según Rampino, dado el contenido en azufre de El Chichón, y puesto que su erupción alcanzó la fuerza suficiente para ascender hasta la estratosfera, quizá tenga una repercusión climatológica comparable a la del monte Agung. La temperatura podría descender en el hemisferio Norte, este año y el próximo, unos 0,5 grados.



Procesos visuales ocultos

La visión, considerada desde siempre un solo sentido, incluye en realidad varios subsistemas especializados cuyas operaciones permanecen ocultas a la conciencia del propio observador

Jeremy M. Wolfe

Quien examine su capacidad para percibir el mundo que le rodea valorará la visión como un solo sentido. Las imágenes formadas en las dos retinas generan una única conciencia de los objetos circundantes: tamaños, formas, colores, texturas y posiciones. Sin embargo, la verdad es muy otra. El sistema visual (los mecanismos cerebrales por los que se procesa la información procedente de los ojos) es, en realidad, un conjunto de subsistemas especializados, que actúan, más o menos independientemente, sobre un subconjunto de datos visuales. Además, algunos de estos subsistemas producen resultados que no son visibles, si bien contribuyen a la función cerebral e incluso a nuestro conocimiento del mundo exterior, sin que el observador sea consciente de los subsistemas propiamente dichos. Es decir, realizan procesos ocultos.

¿Cómo poner de manifiesto un proceso oculto? Mediante el examen, por ejemplo, de las capacidades visuales de personas que han sufrido algún daño en el cerebro. Consideremos la pupila del ojo, que se hace más pequeña ante un aumento de la intensidad de la luz incidente en la retina. Si una persona tiene destruido el córtex visual (la parte del córtex cerebral que recoge en primer lugar la información de los ojos), es perceptualmente ciega, es decir, su conciencia de falta de visión es la misma que la de una persona que hubiera perdido los ojos. Sin embargo, las pupilas siguen modificando normalmente su tamaño en respuesta a las variaciones de luz. Y no sólo eso. Según los descubrimientos de Ernest C. Poppel, Richard Held y Douglas Frost, del Instituto de Tecnología de Massachusetts, cuando a personas que son perceptualmente ciegas, por daños en el córtex visual, se les pide que dirijan su mirada a un punto luminoso, lo hacen sorprendentemente bien. Estas personas manifiestan que no pueden ver el punto, y por ello piensan que están adivinando su posición; sin embargo, aciertan la dirección correcta muchas más veces de las que justificaría el azar.

Personas con lesiones cerebrales evidencian múltiples procesos visuales: unos dañados y otros ocultos, que permanecen en funcionamiento. Lo que estos estudios no pueden confirmar es si algún proceso oculto similar opera también en personas con el sistema visual sin alterar. Para ello es necesario diseñar estrategias experimentales especiales. Describiré a continuación tres conjuntos de experimentos, cada uno de los cuales pone de manifiesto un proceso visual que en personas normales está oculto a la introspección. A este respecto, todos somos como las personas con el cerebro dañado, que pueden mirar hacia un punto luminoso aunque estén incapacitados para percibirlo. Tampoco tenemos conciencia de la capacidad de algunas partes del sistema visual, aun cuando nuestro comportamiento se aproveche a menudo de los resultados del funcionamiento de estos subsistemas.

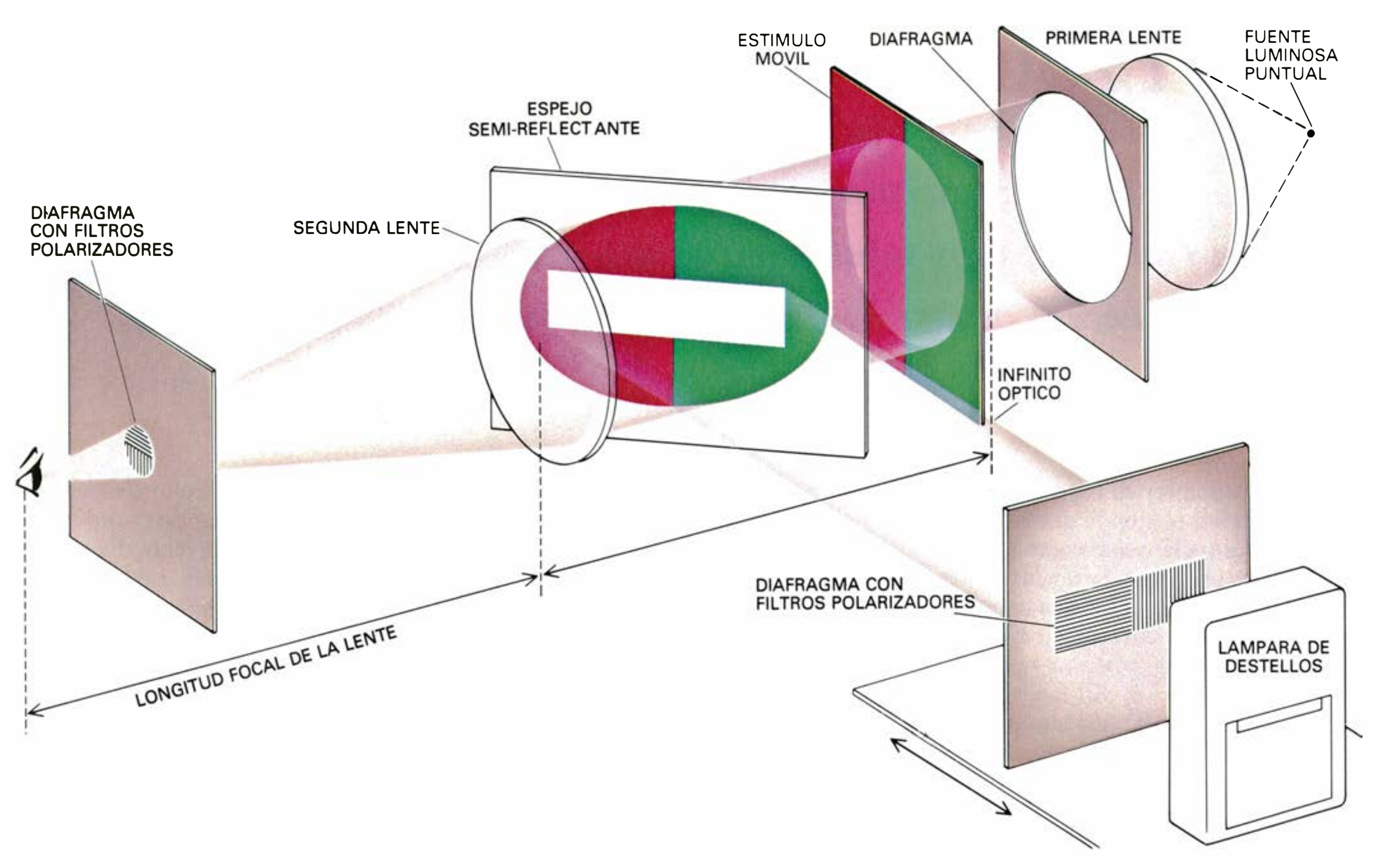
Una de las funciones del sistema visual es controlar los músculos que enfocan el ojo sobre los objetos, situados a diferentes distancias, mediante cambios de la forma de la lente del ojo, el cristalino. Cuanto más cerca está el objeto, más esférico se hace el cristalino. Este proceso se llama acomodación visual. Tiene uno la impresión de que puede acomodarse a cualquier objeto visible y, por tanto, es natural suponer que acomodación y percepción visual tienen acceso al mismo conjunto de estímulos.

¿Es ello cierto? El estado de acomodación del ojo de un observador que mira a un estímulo situado a determinada distancia puede medirse por diversos métodos experimentales. En uno, el observador mira un estímulo a través de filtros polarizadores y, mediante un flash colocado detrás de una rendija, se superpone brevemente sobre el estímulo una línea brillante horizontal. Los filtros no influyen en el estímulo; tienen por misión conseguir que la luz de la mitad izquierda de la línea brillante entre sólo por la parte superior del cristalino del ojo del observador y la luz de la mitad derecha entre sólo por la parte inferior.

Supongamos que un observador está mirando el estímulo, pero acomodando a una distancia superior o inferior a la distancia a la que se encuentra la rendija. Si ésta se ilumina en ese instante, las dos mitades no estarán alineadas en la retina, y el observador las verá desplazadas. Así, el experimentador sólo necesita que el observador mire el estímulo y, mientras se ilumina repetidamente la rendija, preguntarle si ve las dos mitades alineadas. La rendija se desplaza para modificar la distancia hasta que el observador considera que las dos mitades están alineadas. Cuando así le parece, la distancia de acomodación es la distancia entre la rendija y el ojo del observador.

El cristalino se endurece con los años, por lo que es conveniente estu-

1. SE LLAMA ESTIMULO ISOLUMINANTE a la imagen cuyos bordes están definidos por cambios de color y no por cambios de claridad. El estímulo de la figura es imperfecto, ya que las partes azules y las verdes no tienen exactamente la misma claridad; ofrecen la igualdad máxima que permite la impresión en papel. Además, es manifiesto el cambio de claridad entre la página y el entorno, así como el hecho de que el lector mantiene la revista a la distancia de lectura. Si en condiciones de laboratorio se eliminan estas claves, un observador es incapaz de enfocar los bordes de un estímulo isoluminante. Esta deficiencia hace que sea difícil reconocer hasta una cara familiar. El experimento indica que el proceso cerebral responsable de la acomodación visual (el enfoque de los ojos) no puede "ver" el color. Es un proceso oculto distinto del proceso perceptual. La figura muestra a Groucho Marx en la película *Plumas de caballo*.



2. SISTEMA EXPERIMENTAL diseñado para el estudio de la capacidad de acomodación a un borde de un estímulo isoluminante. La primera lente, situada en el ángulo superior derecho, colima un haz de luz. Este haz atraviesa un diafragma y, luego, el estímulo, que tiene un borde vertical definido exclusivamente por los colores rojo y verde. Una segunda lente dirige el haz hacia el ojo del observador. Mediante un dispositivo adicional, formado por una lámpara de destellos, una rendija móvil y un espejo semirreflectante, se superpone brevemente una línea horizontal sobre el estímulo insoluminante.

RETINA
CRISTALINO
a b c
FILTROS POLARIZADORES

a

b

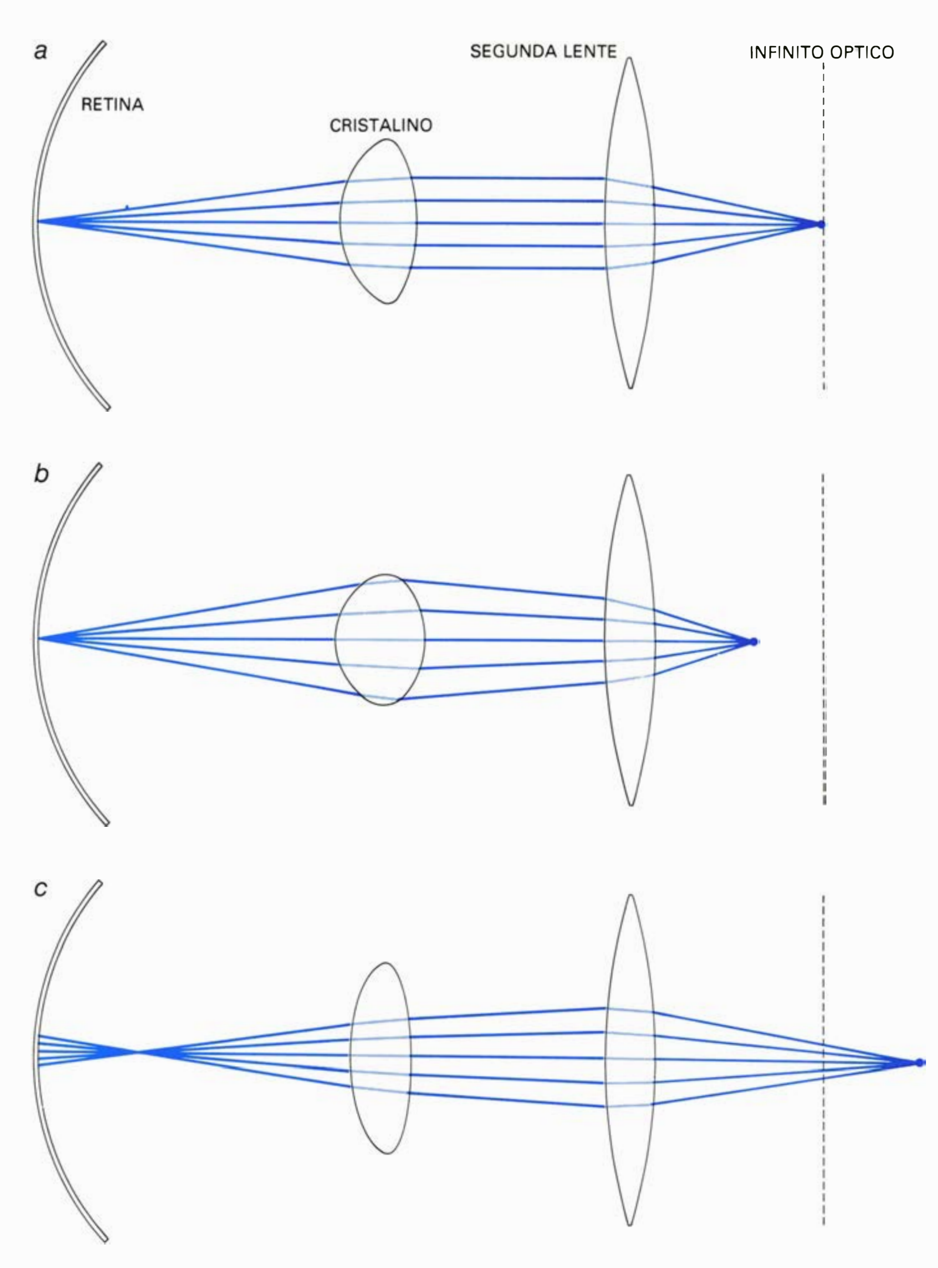
c

3. PARA MEDIR LA ACOMODACION se atiende a la evolución de la línea horizontal. Un conjunto de filtros polarizadores aseguran que la luz del lado izquierdo de la rendija entre sólo por la parte superior del cristalino del observador y la luz del lado derecho por la parte inferior. Los dos lados de la rendija se alinean en la retina (*b*) sólo cuando el sujeto experimental acomoda la vista a la distancia de la rendija; en otro caso, aparecen desplazadas.

diar la acomodación en personas que no hayan alcanzado los 40 años. A los observadores se les presenta, en el laboratorio, estímulos que no cambian de brillo ni tamaño, modificándose sólo la distancia al observador. Así se eliminan todas las claves perceptuales de la distancia al estímulo, excepto la de que, para cierto estado de acomodación, el estímulo queda enfocado. Bajo estas condiciones, la acomodación de un observador típico es perfecta en un 90 por ciento. (Si fuera perfecta del todo, un estímulo situado a 25 centímetros induciría un estado de acomodación del ojo a 25 centímetros, por ejemplo.)

¿Cómo responde el ojo cuando el estímulo no tiene ninguna característica que le induzca a enfocarlo? Tendríamos un estímulo de esta clase en una pantalla uniforme que rodeara al observador, como si se encontrara dentro de una pelota gigante de ping-pong. Un estímulo similar sería una pantalla pequeña y uniforme, vista a través de una lente de manera que la pantalla parezca demasiado próxima (en distancia óptica o aparente) para que el ojo pueda enfocar los bordes. Otros estímulos serían un cielo sin estrellas, o la oscuridad absoluta. Herschel W. Leibowitz y D. Alfred Owens, de la Universidad estatal de Pennsylvania, han observado que en cualquiera de estas circunstancias el cristalino adquiere un estado de reposo para una distancia específica bastante estable, que se llama foco en oscuridad. Cada observador tiene un foco en oscuridad característico, que está, normalmente, a un metro o a una distancia equivalente a la longitud del brazo.

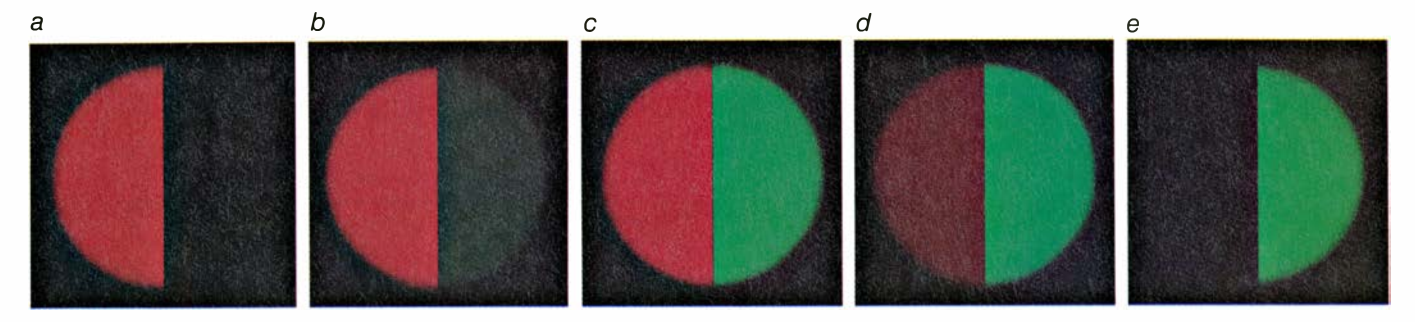
Lo mismo ocurre cuando el estímulo está constituido por una serie de líneas blancas y negras lo suficientemente juntas para que den la impresión de un campo gris. El experimentador puede colocar el estímulo a diversas distancias del ojo del observador mediante una lente. Cuando el estímulo se halla a una distancia óptica de 25 centímetros, el cristalino del observador adquiere su estado de reposo y, cuando el estímulo se desplaza a una distancia de un metro, la acomodación no se modifica. La conclusión es que la acomodación induce cambios en la forma del cristalino para mantener un objeto enfocado sólo cuando al sistema acomodativo se le presenta algo que pueda "ver".



4. DISTANCIA OPTICA de un estímulo es la distancia aparente proporcionada por la segunda lente del sistema experimental, para una posición dada del estímulo. Si el estímulo está en el plano focal de la lente (*a*) se verá en el infinito óptico, es decir, los rayos procedentes de cada punto del estímulo entrarán en el ojo paralelos, como si vinieran de un punto infinitamente alejado. El cristalino adoptará la curvatura adecuada para concentrar los rayos en un punto de la retina. Si el estímulo está más cerca que el infinito óptico (*b*), los rayos divergerán al acercarse al ojo. El cristalino adoptará entonces una forma más esférica para llevar los rayos a foco. Si el estímulo está más lejos que el infinito óptico (*c*), los rayos convergerán al acercarse al ojo y el cristalino será incapaz de reducir su curvatura hasta el grado requerido, enfocando los rayos delante de la retina. Por tanto, un estímulo más lejano que el infinito óptico no se puede enfocar.

Sabiendo que el sistema acomodativo era ciego a ciertos aspectos del mundo visual, Owens y el autor se preguntaron si el sistema visual podía "ver" colores. Para ello medimos la acomodación de los observadores mientras éstos seguían un estímulo simple: un campo circular dividido verticalmente por la mitad para que incluyera sólo un borde vertical. Un sistema óptico del tipo conocido por visión maxwelliana aseguraba que el perímetro del círculo quedara más allá del infinito óptico, es decir, los rayos procedentes de los puntos del perímetro se hacían entrar en el ojo de tal manera que no podían enfocarse nunca, independientemente de la forma que adquiriese el cristalino. Con ello, el único detalle del estímulo donde el ojo podía acomodar era el borde vertical entre las dos mitades del campo. El sistema óptico estaba diseñado para que, además, mantuviera el tamaño de la imagen del estímulo en la retina cuando se cambiaba la distancia óptica.

El borde puede generarse por una diferencia de claridad o de color entre las dos mitades. En el mundo real, la mayoría de los contornos presentan diferencias en ambos factores. Escogimos una mitad roja y la otra mitad verde; variamos la claridad de una mitad a otra. Creábamos así estímulos que iban desde un borde que separaba rojo y ne-



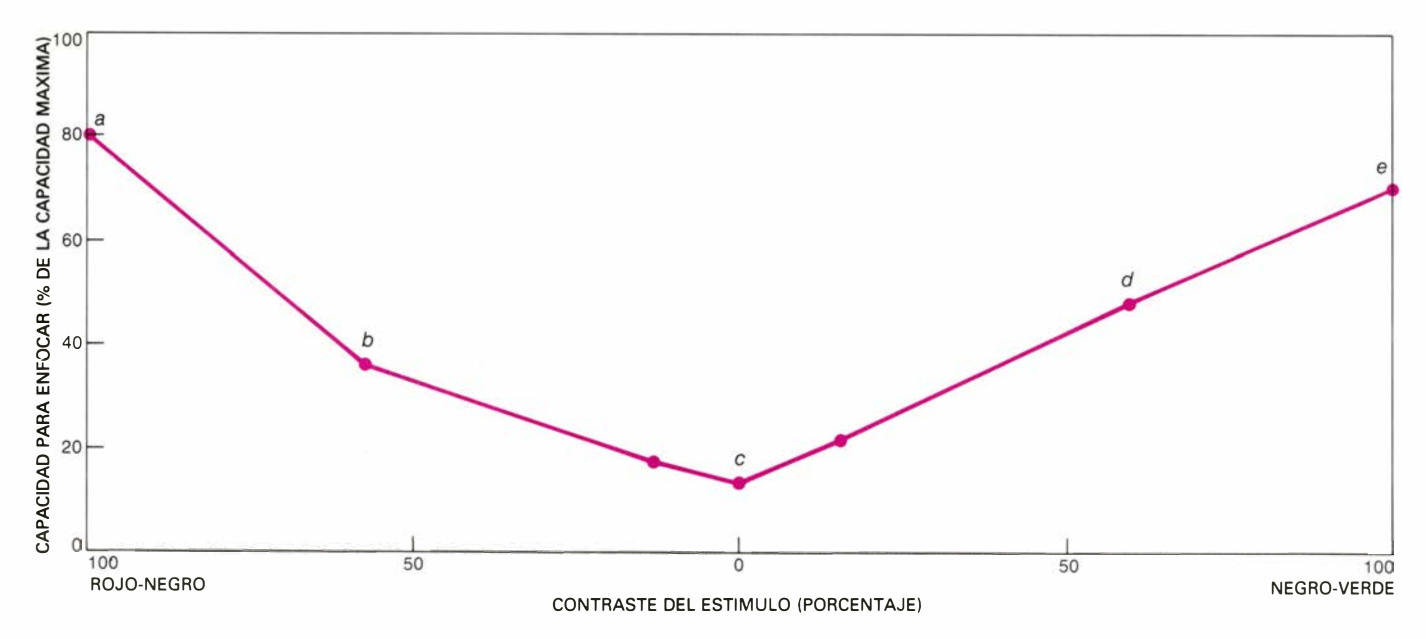
5. LOS ESTIMULOS PARA EL EXPERIMENTO sobre la capacidad de acomodación van desde un borde entre rojo y negro (*a*) hasta un borde entre negro y verde (*b*); en ambos, el contraste de claridad es del 100 por ciento. El estímulo intermedio (*c*) es isoluminante. El perímetro circular del estímulo está definido por un diagrama colocado más lejos que el infinito óptico en el sistema experimental, por lo que el observador no lo puede enfocar.

gro hasta un borde entre negro y verde. En ambos casos extremos, el contraste del borde entre el color y el negro era del 100 por ciento; en la situación intermedia, la mitad roja y la verde tenían la misma claridad. Constituía ello un estímulo isoluminante en el cual el borde estaba definido sólo por el color, siendo nulo el contraste en claridad.

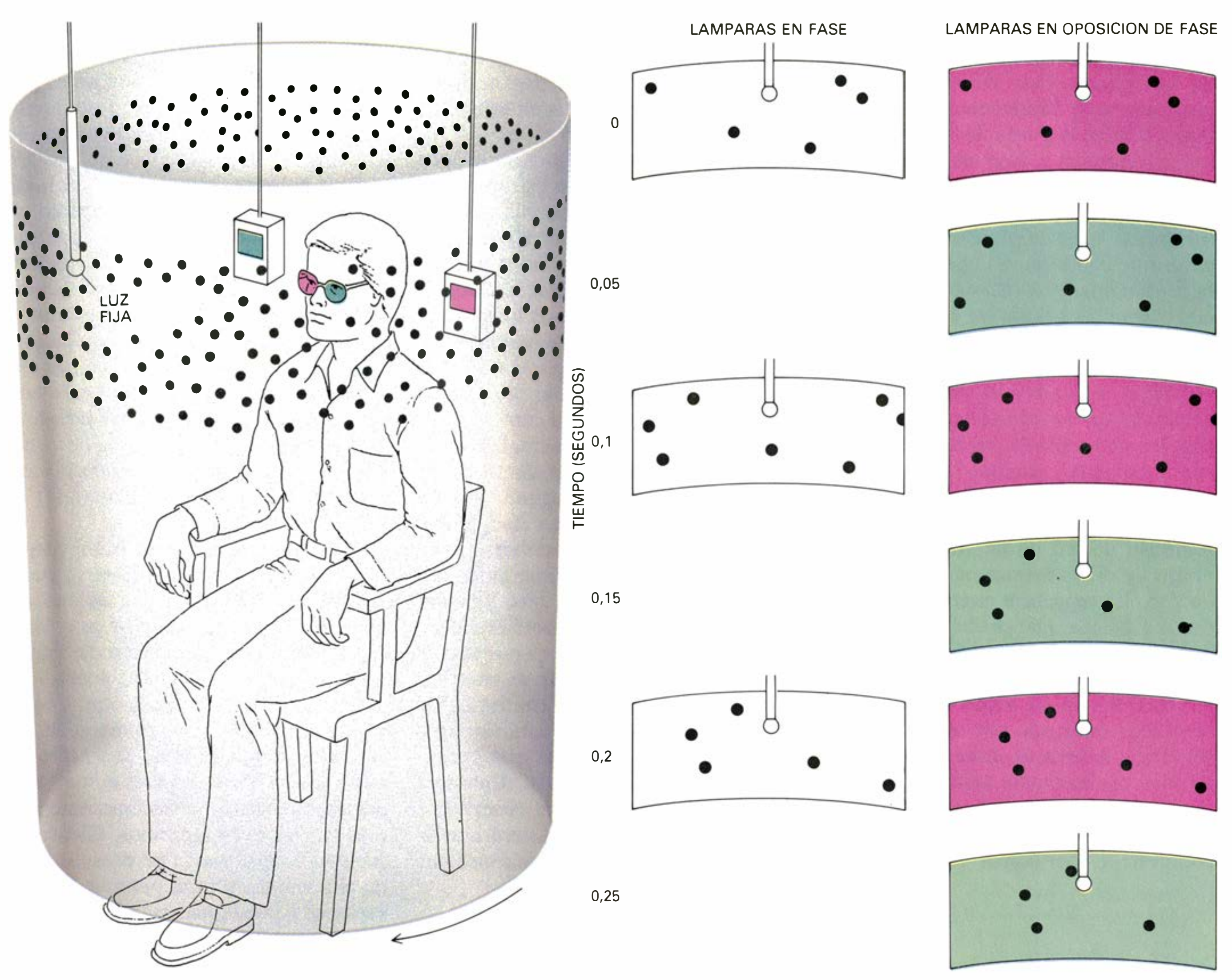
Al estudiar la capacidad de acomodación a estos estímulos observamos que disminuía a medida que lo hacía el contraste de claridad. Nuestros observadores reconocían con facilidad que el estímulo isoluminante era un borde entre rojo y verde, pero no podían enfocarlo. En otro experimento, los observadores ponían sus ojos sobre una letra *E* negra sobre fondo blanco. Los sujetos acomodaban perfectamente a distancias ópticas entre el infinito y 22 centímetros. Les pedíamos que mirasen luego bordes entre rojo y negro o entre negro y verde. Su acomodación era de un 80 por ciento, comparada con la que habían conseguido con la letra *E*. Por último, les mostrábamos el borde rojo-verde sin contraste de claridad; la acomodación sólo alcanzaba un 19 por ciento. Cuando repetimos el experimento con otros pares de colores, rojo y naranja o azul y verde, la actuación del observador era tan pobre o peor.

Análogamente a cualquier sistema óptico, el ojo tiene aberración cromática, esto es, los distintos colores quedan enfocados a distancias focales ligeramente diferentes. La aberración tiende a desplazar la imagen retiniana de un color respecto a la del otro, lo que genera un contorno brillante, u oscuro, según que las imágenes se separen o se superpongan. Por ello sorprende cuán difícil resulta crear un estímulo que no tenga ninguna otra clave de la presencia de un borde que la diferencia de color. Sospechamos que la actuación de nuestros observadores habría sido aún peor si se hubiese eliminado la aberración cromática. La conclusión parece clara: la acomodación es ciega al color, o por lo menos "ve" poco los colores. Es un subsistema visual marcadamente independiente de la percepción visual. Es, en definitiva, una característica de la visión que no podría haberse encontrado sólo por introspección.

Una segunda estrategia para explorar la división del sistema visual en subsistemas es considerar las formas en que las diferentes partes del sistema realizan una determinada función visual. La visión binocular, por ejemplo. Una persona normal se sirve de los dos ojos; el cerebro hace lo necesario para integrar los datos de ambos. Sin embargo, parece claro que el cerebro hace algo más que una simple combinación de las señales de los dos ojos en un solo proceso binocular que conduce a la percepción visual. Existen diversos mecanismos específicos que combinan las señales a su manera con el fin de satisfacer sus propias necesidades particula-



6. CAPACIDAD PARA ENFOCAR el estímulo. Depende del contraste de claridad. En la figura se representa el comportamiento de cuatro observadores como porcentaje de la capacidad lograda acomodando la visión a letras negras de un test de optotipos. La capacidad para enfocar disminuye a medida que disminuye el contraste, siendo mínima (19 por ciento de la capacidad máxima) para el estímulo insoluminante, que corresponde a *c* en la figura 5.



7. EXPERIMENTO SOBRE VECTACION (sensación de automovimiento generada por el movimiento del mundo visual). Los observadores se sentaban dentro de un cilindro giratorio con la superficie interior pintada de blanco con manchas circulares negras distribuidas al azar. Con dos lámparas estroboscópicas se iluminaba la superficie interior con destellos de luz verde y roja. Unas gafas limitaban la entrada de cada color a un solo ojo. Si las lámparas se encendían en fase (es decir, en sincronía) a una frecuencia de 10 destellos por segundo, un observador con un solo ojo abierto veía lo mismo que un observador con los dos ojos abiertos: una película de círculos negros en movimiento a 10 imágenes por segundo. Si las lámparas estaban en oposición de fase, un observador con un solo ojo abierto veía esa misma película, pero un observador con los dos ojos abiertos veía una película más. Resultaba así una película interocular a 20 imágenes por segundo. Los fotogramas se veían, alternativamente, con los dos ojos. Una pequeña luz fija servía de punto de referencia.

res. De igual forma que por introspección no se puede llegar a saber que la acomodación es ciega al color, tampoco por introspección pueden ponerse de manifiesto estos procesos múltiples. Sin embargo, tales procesos existen y pueden aflorar en los experimentos.

Consideremos una ilusión visual que todos hemos experimentado alguna que otra vez. Estamos sentados en un departamento del tren, esperando que arranque. Por la ventana vemos otro tren parado en la vía de al lado. La imagen del otro tren comienza a deslizarse hacia atrás, y experimentamos la innegable impresión de que el nuestro avanza. Pronto advertimos que es el ferrocarril de al lado el que se mueve; el nuestro sigue detenido. El movimiento del otro tren nos indujo erróneamente a creer que nos desplazábamos; hemos experimentado una ilusión de movimiento causada por mera estimulación visual. Tal ilusión se conoce por vectación.

Muchos investigadores han intentado descifrar los mecanismos cerebrales responsables de la vectación. Held y yo nos interesamos en una cuestión algo diferente. Queríamos saber qué es lo que el proceso de vectación podía "ver". En particular queríamos saber si el proceso se valía de la capacidad de mirar el mundo a través de dos ojos. Quizá parezca que para dar con la respuesta bastará comparar la vectación experimentada con los dos ojos abiertos con la experimentada con un solo ojo abierto. Sin embargo, la sensación de vectación suele ser máxima con un solo ojo abierto; la ilusión de automovimiento no puede ser más pronunciada. Es necesario, pues, crear un estímulo cuyo efecto binocular sea, en principio, mayor que su efecto monocular. Tal estímulo debe tener una componente puramente binocular, es decir, debe tener algo que sea invisible a cada ojo por separado.

Una componente que cumpla estos requisitos se llama estímulo ciclópeo, de Cíclope, el personaje homérico de un solo ojo con quien topó Ulises. El nombre es adecuado, puesto que un estímulo ciclópeo se hace evidente sólo cuando el cerebro combina las señales de los dos ojos. Es decir, el cerebro debe actuar como un único ojo, ojo ciclópeo. Un ejemplo de estímulo ciclópeo

es la pequeña diferencia en la posición de la imagen de un objeto en cada una de las retinas. Esta diferencia puede mostrarse muy fácilmente: extienda un brazo y mire la punta de un dedo con un ojo cerrado y después con el otro ojo. Apreciará que las dos observaciones son ligeramente diferentes. El cerebro utiliza estas diferencias en los procesos que conducen a la percepción de la profundidad; obviamente, para que ello ocurra hace falta que el cerebro recoja datos de los dos ojos.

En nuestro estímulo ciclópeo para la vectación utilizamos el muy conocido hecho de que los objetos que se muestran en una película parecen moverse continuamente, aunque lo que se proyecta sea una sucesión de fotogramas fijos. Nuestros observadores se sentaban dentro de un cilindro de un metro de diámetro por metro y medio de alto. La superficie interior del cilindro era blanca, manchada con círculos negros de unos 2,5 centímetros de diámetro distribuidos al azar. El cilindro giraba sobre su eje a una velocidad de 30 grados de ángulo por segundo. Si la superficie interior se iluminaba con luz normal, el observador manifestaba sensación de giro en sentido contrario. En nuestro caso iluminamos la superficie con destellos periódicos de una lámpara estroboscópica. Cada destello producía el equivalente a un fotograma de una película de puntos negros; el observador manifestaba la misma sensación de vectación.

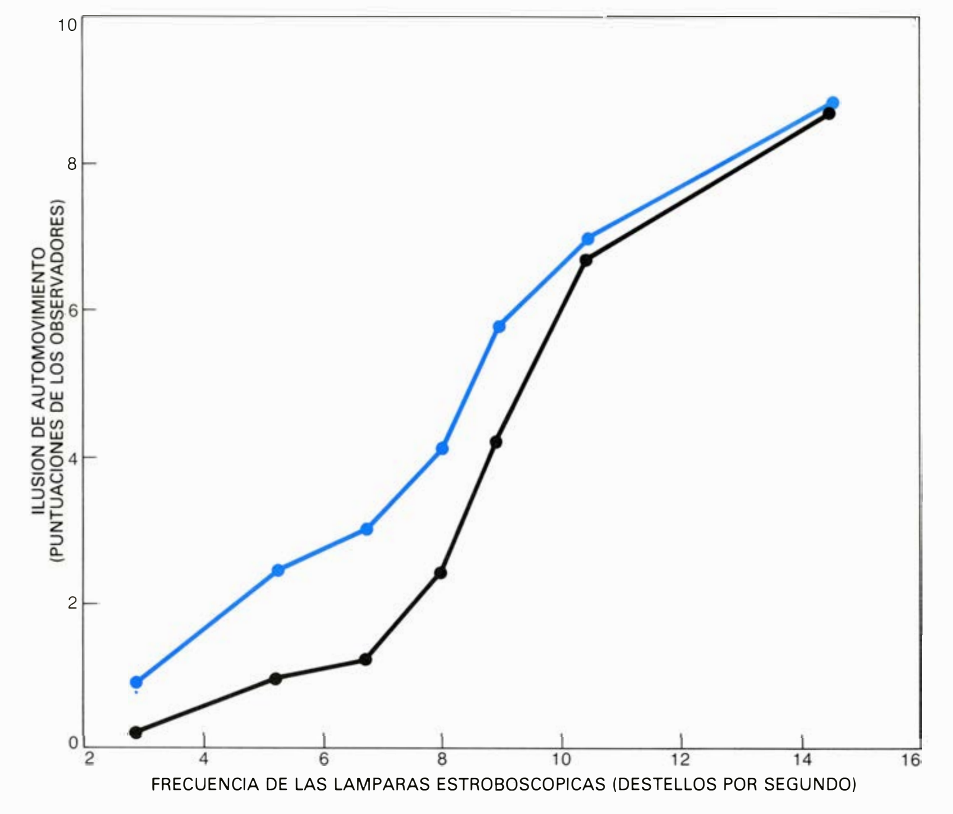
Para producir un estímulo ciclópeo hacían falta dos lámparas estroboscópicas, una con un filtro rojo y la otra con un filtro verde. Los observadores utilizaban unas gafas con un filtro rojo en un ojo y un filtro verde en el otro. El filtro verde no dejaba pasar luz roja, ni el filtro rojo luz verde. De esta manera, la luz de una de las lámparas estroboscópicas se veía con un solo ojo y la luz de la otra lámpara con el otro.

Supongamos que las lámparas se encienden a una frecuencia de 10 hertz, esto es, 10 veces por segundo. Si las lámparas se encienden en fase, o sincronía, el observador no debería notar diferencias importantes entre mirar con un solo ojo o con los dos ojos. En cualquiera de esos dos casos verá de hecho una película, a 10 imágenes por segundo, de círculos negros que se mueven. Supongamos que, por el contrario, las lámparas estroboscópicas están en oposición de fase: cuando una está encendida la otra está apagada. Entonces existe una diferencia: si el observador tiene un solo ojo abierto, verá nuevamente una película de 10 imágenes por segundo, mientras que con los dos ojos abiertos verá dos. En conjunto contemplará una película de 20 imágenes por segundo, consistente en fotogramas presentados alternativamente a un ojo y al otro.

Esta película interocular a 20 hertz constituye, por tanto, un estímulo ciclópeo: no puede verse con cada ojo por separado y causa sensación de movimiento. La pregunta es si el sistema de vectación puede "ver" este estímulo. Si así fuera, el caso en el que las dos lámparas estén en oposición de fase debería producir mayor sensación de vectación que el caso en que las lámparas se hallen en fase. Si el sistema de vectación no "ve" el estímulo, en los dos casos se tendría el mismo resultado.

Una vez dispuesto el aparato, preparamos a los observadores para el experimento. En concreto les enseñamos un método conocido en psicología experimental por estimación de magnitudes o método de puntuaciones. Les pedíamos que puntuaran con 10 una fuerte sensación de automovimiento y con cero la ausencia de sensación de movimiento. Puntuaciones entre 0 y 10 corresponderían a sensaciones intermedias entre los dos extremos. Contrariamente a lo que pueda parecer, se tiene mucha habilidad para puntuar las sensaciones y puede hacerse repetidamente y con fiabilidad, como lo han demostrado 25 años de investigación en psicología experimental, en especial el trabajo de S. Smith Stevens, de la Universidad de Harvard.

Quedaba aún un punto importante: si la sensación de vectación fuera de 10 cuando las lámparas estaban en fase, no habría ninguna diferencia entre los resultados de los dos experimentos. Cuando los fotogramas de una película se proyectan con frecuencias progresivamente decrecientes, la ilusión de movimiento pierde intensidad. Análogamente, la frecuencia de los destellos es muy importante para la sensación de automovimiento. Encontramos que cualquier frecuencia por encima de 2,5 hertz producía un cierto grado de vectación, pero sólo frecuencias por encima de los 15 hertz repetían una puntuación de 10 en las anotaciones de nuestros observadores. Así pues, las frecuencias comprendidas entre 2,5 hertz y 15 hertz serían las adecuadas para nuestros propósitos. El estímulo ciclópeo creado (al doble de la frecuencia de destello) encendiendo las lámparas en oposición de fase debería producir en principio mayor sensación de vectación



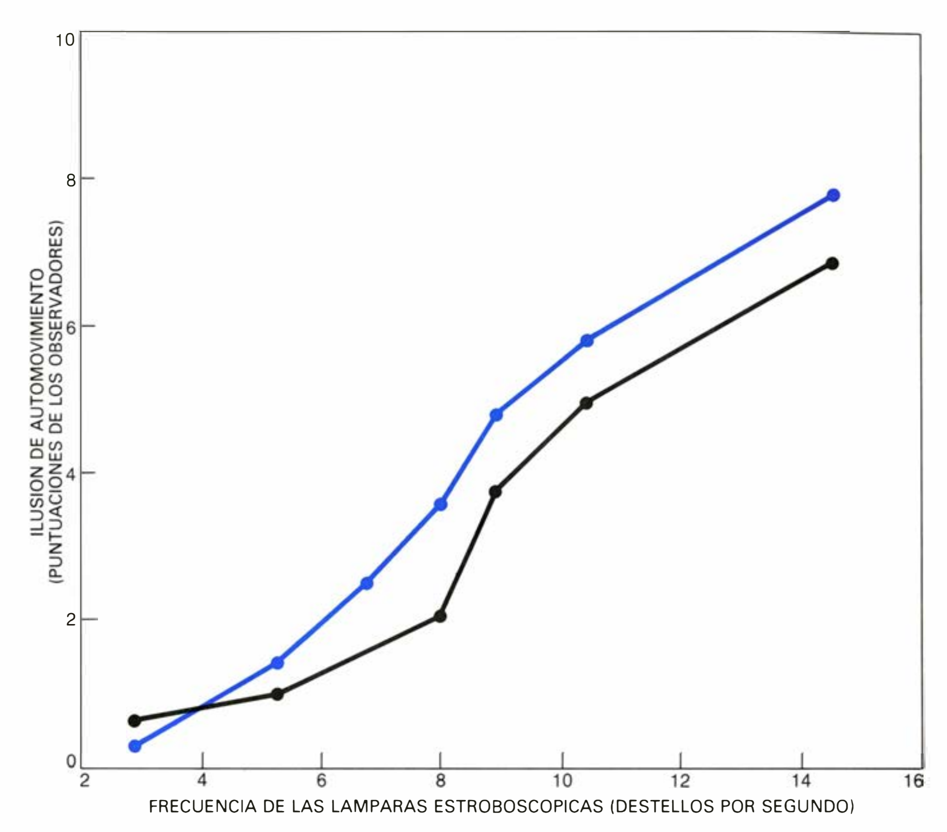
8. LA SENSACION DE VECTACION cuando las lámparas están en oposición de fase (*color*) es mayor que cuando se hallan en fase (*negro*). Demuestra ello que el proceso cerebral responsable de la vectación puede "ver" una película estimulando los ojos en alternancia; es un proceso binocular. En la figura se representan los datos de seis observadores que puntuaron de 0 a 10 la magnitud de la sensación percibida.

que el estímulo obtenido dejando las lámparas en fase.

Los resultados fueron concluyentes. A todas las frecuencias de destello entre 2,5 hertz y 15 hertz, las lámparas en oposición de fase obtenían las mayores puntuaciones. Evidentemente, el sistema de vectación puede "ver" el estímulo puramente binocular. ¿Significa ello que el cerebro tiene múltiples procesos binoculares? No del todo. Como he mencionado antes, la pequeña diferencia en la posición de las imágenes de un objeto en las dos retinas es una clave ciclópea para la profundidad tridimensional. Habrá que pensar que el mecanismo binocular que sirve para este aspecto de la percepción visual sirve también para la contribución binocular a la vectación.

Pero hay individuos incapaces de percibir la profundidad por medio de un estímulo ciclópeo. Así, por ejemplo, estas personas no tienen ilusión de relieve cuando miran una imagen en un estereoscopio o cuando ven una película tridimensional. No tienen visión estereoscópica, es decir, son ciegos al relieve. Otros pierden esta cualidad a causa de defectos desarrollados en la infancia, que afectaron la capacidad de alinear los ojos. La falta de visión de relieve es algo menos frecuente que la falta de visión de color y, al parecer, se circunscribe a un pequeño porcentaje de la población. Una persona sin visión de relieve puede percibirlo a pesar de todo, pero a través de claves como el tamaño aparente de objetos familiares o por el hecho de que algunos objetos del campo visual tapan a otros que están detrás. Estas claves monoculares del relieve son muy eficaces; incluso muchas personas sin visión estereoscópica no lo saben hasta después de hacerse algún examen.

Held y yo solicitamos de cuatro sujetos, con probada falta de visión de relieve, que hicieran de observadores en nuestras prácticas de vectación. Sus puntuaciones fueron muy parecidas a las de los observadores normales. También percibieron un fuerte aumento en la sensación de automovimiento con el estímulo puramente binocular. Es decir, hay personas con un mecanismo binocular defectuoso que, sin embargo, se comportan como normales en experimentos en los que se requiere que el cerebro utilice un mecanismo binocular. El mecanismo defectuoso no puede ser el único mecanismo binocular del cerebro. Tiene que haber, por lo menos, dos mecanismos binoculares, uno para la visión estereoscópica de la profundidad y otro implicado en la producción de la vectación. La experimentación demuestra, una vez más, que los datos visuales alimentan más de un sistema de procesamiento.



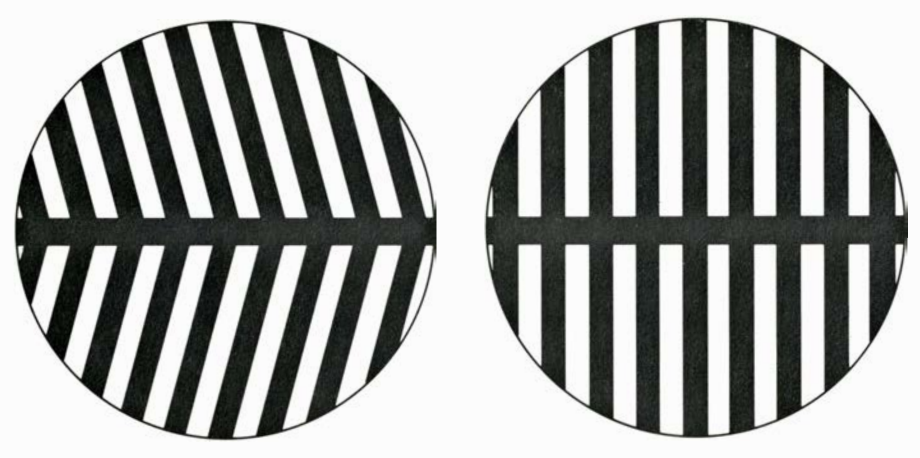
9. PUNTUACIONES DE OBSERVADORES SIN VISION ESTEREOSCOPICA. Muestran que éstos tienen mayor sensación de vectación cuando las lámparas están en oposición de fase; pero carecen del proceso visual binocular que compara las imágenes de los dos ojos y que conduce a la percepción de profundidad tridimensional. El proceso responsable de la vectación debe ser, pues, un proceso binocular diferente. Los datos corresponden a las puntuaciones de cuatro observadores sin visión estereoscópica.

Los dos procesos visuales que acabo de describir (el proceso ciego al color que controla la acomodación y el proceso binocular que contribuye a la vectación) no parece que desempeñen un papel directo en la percepción visual, pero existen otros procesos visuales ocultos que sí toman parte en la percepción.

A modo de ejemplo, volveré a los procesos visuales binoculares. Nuestros experimentos con la vectación revelaban la existencia de dos procesos binoculares. Los experimentos que describiré a continuación muestran más; la verdad es que aparece una extensa gama de procesos. Entre los dos ojos y una única percepción visual del mundo median procesos que pueden "ver" a través de un ojo, procesos que pueden "ver" a través de uno cualquiera de los ojos y, sorprendentemente, un proceso binocular puro: un proceso que puede "ver" sólo a través de los dos ojos. De hecho sólo puede "ver" cuando ambos ojos están contemplando el mismo estímulo. Si el sistema visual tuviera sólo el proceso binocular puro, sería incapaz de ver con un ojo cerrado. Por ejemplo, sería incapaz de ver si miráramos desde detrás de un árbol, y un ojo viera sólo corteza mientras el otro contemplara una casa.

La capacidad de percibir un proceso puramente binocular, como el estímulo ciclópeo que utilizamos en los experimentos de vectación, no asegura que el cerebro tenga un proceso exclusivamente binocular. El estímulo ciclópeo podía haber sido "visto" por un proceso visual que aceptara señales de ambos ojos y también de cada uno de los dos ojos por separado. ¿Cómo se sabe, entonces, si tenemos en efecto un estímulo puramente binocular? Nuevamente, recurrimos a métodos indirectos, en este caso basados en una distorsión temporal de la visión llamada post-efecto de inclinación.

Mire el círculo derecho de barras de la figura 10. La mitad superior debe verse alineada con la mitad inferior. Fíjese ahora en las barras en ángulo de la izquierda. Deje que sus ojos se paseen a lo largo de la barra negra horizontal que cruza por el centro de las barras en ángulo durante uno o dos minutos y rápidamente mire el dibujo de la



10. POST-EFECTO DE INCLINACION. Sirve esta ilusión para el estudio de los procesos visuales ocultos de las vías que conducen a la percepción visual. Para experimentar esta ilusión desplace la vista de un lado a otro por la línea negra que cruza las barras en ángulo del círculo izquierdo de la ilustración. Observará luego que, por unos momentos, las barras de la derecha no son colineales. El post-efecto se mide ajustando las dos mitades del dibujo de la derecha; se pide a los observadores que ajusten las barras para que sean colineales inmediatamente después de haber mirado la figura de la izquierda.

derecha. Debería ver que las dos mitades ya no están alineadas, sino inclinadas en sentido opuesto a la inclinación de las barras de la izquierda.

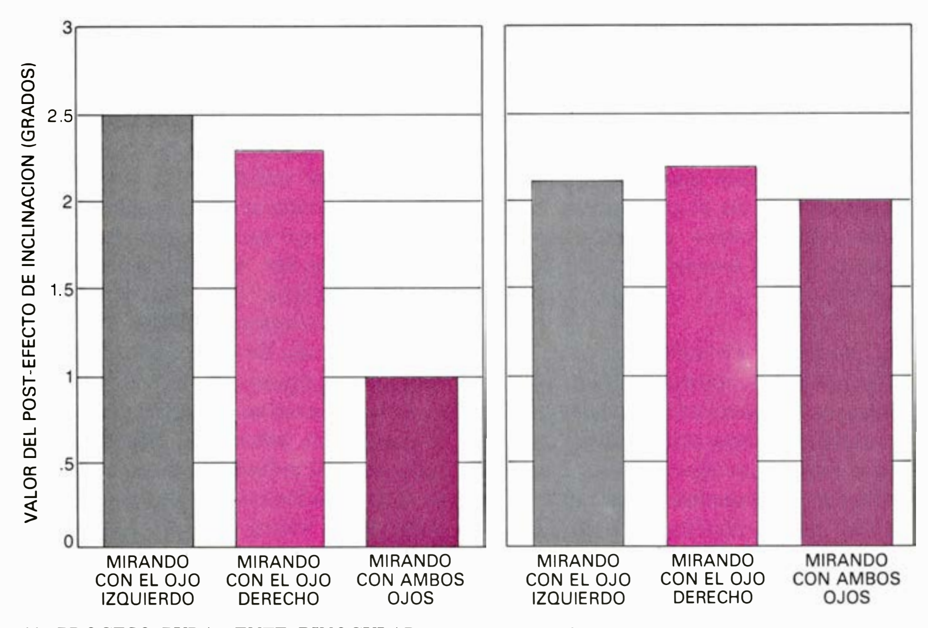
Este es el post-efecto de inclinación. Se puede medir haciendo que las dos mitades del dibujo de la derecha sean ajustables y pidiendo a los observadores que las alineen. Antes de mirar las barras en ángulo, las alineaciones se acercaban prácticamente a la colinealidad. Después de mirar las barras en ángulo, las alineaciones quedaban sistemáticamente desplazadas unos dos grados.

Supongamos que una persona mira las barras en ángulo con el ojo derecho solamente y manipula el dibujo ajustable mirando nada más que con el izquierdo abierto. Se detecta el post-efecto a pesar de que el ojo izquierdo no ha visto en ningún momento las barras en ángulo, esto es, existe transferencia inter-ocular. Ello supone una prueba de la existencia de un proceso binocular, pero no exclusivamente binocular. Después de todo, el proceso binocular se activaba cuando el ojo derecho observaba las barras en ángulo y permanecía activado cuando el izquierdo alineaba el dibujo ajustable. Aparentemente, el proceso puede responder a cualquiera de los ojos.

El post-efecto es menos intenso cuando lo mide el ojo que no ha estado expuesto a las barras en ángulo que cuando lo mide el que sí ha estado expuesto. Se transfiere sólo del 70 al 80 por ciento del post-efecto. Randolph Blake y sus colegas, de la Universidad de Northwestern, concluyeron de ello que cuando el post-efecto de inclinación se examina con un ojo hay al menos dos procesos implicados. Uno de ellos es un proceso binocular; el otro, monocular. Así, la exposición del ojo derecho a las barras inclinadas activa ambos procesos, el binocular y el monocular del ojo derecho. El examen del ojo derecho activa de nuevo ambos procesos. El resultado es un fuerte post-efecto de inclinación. Examinemos ahora el ojo izquierdo. El proceso binocular que estaba expuesto al estímulo se activa también ahora. El proceso monocular del ojo izquierdo opera, a su vez; pero este ojo no vio nunca las barras en ángulo. Su respuesta diluye la magnitud del post-efecto.

Así, por medio del post-efecto de inclinación, Blake demostró la existencia de ambos procesos, el monocular y el binocular. Held y el autor se sirvieron del post-efecto para demostrar que también existe un proceso exclusivamente binocular. Hacíamos que los observadores mirasen las barras en ángulo con cada ojo alternativamente: un minuto con el ojo izquierdo y un minuto con el ojo derecho, otro minuto con el ojo izquierdo de nuevo y otro con el ojo derecho. Para medir el post-efecto, abrían el ojo derecho, o el ojo izquierdo, o ambos.

Pensábamos que la observación monocular alternada del estímulo excitaría todos los procesos que se alimentan con señales del ojo izquierdo, todos los procesos que se alimentan con señales del ojo derecho y todos los procesos que se alimentan con las señales de uno u otro ojo. Si éstos fueran los únicos procesos visuales, constituirían el sistema visual completo; no habría, pues, diferencia entre tener un ojo abierto o los dos cuando el observador manipulara las barras ajustables: el post-efecto de inclinación sería el mismo. Si, por el contrario, existiese un proceso visual que fuera activo sólo cuando los dos ojos se



11. PROCESO PURAMENTE BINOCULAR, puesto de manifiesto comparando los resultados de los experimentos que utilizan el post-efecto de inclinación. En uno de los experimentos (*izquierda*) los observadores miraban las barras en ángulo alternando los dos ojos; se activaban así todos los procesos visuales que toman señales de un ojo u otro. Tal estímulo no activaba un proceso binocular puro, es decir, un proceso que tomara señales, a la vez, de ambos ojos. Los observadores manipulaban luego las dos mitades ajustables mirando con un ojo, o bien con los dos. El post-efecto binocular era menor que el monocular. El segundo experimento (*derecha*) difería del anterior en que los observadores atendían a la figura de barras en ángulo con los dos ojos a la vez. El post-efecto binocular y el monocular eran entonces iguales. La observación binocular del estímulo en este segundo experimento activó el proceso binocular puro.

estimulan simultáneamente, los resultados serían diferentes. La observación monocular alternada del estímulo no excitaría este proceso, pero si ambos ojos estuvieran abiertos durante la medida del post-efecto, el proceso se activaría y disminuiría el post-efecto.

Eso fue lo que encontramos. El post-efecto era mucho menor al ajustar las barras con los dos ojos abiertos que con un ojo cerrado. En un experimento complementario, les hacíamos mirar a los observadores las barras en ángulo con los dos ojos abiertos. Esperábamos que así se excitarían todos los procesos visuales; estábamos en lo cierto: el post-efecto era ahora prácticamente igual, tuviera el observador uno o los dos ojos abiertos.

En ambos experimentos, el ojo izquierdo y el ojo derecho estaban expuestos al estímulo de barras en ángulo. La única diferencia era que en el primer experimento los observadores nunca veían el estímulo con los dos ojos abiertos a la vez, mientras que en el segundo experimento sí. Había un proceso que no se excitaba en el primer experimento y que se excitaba después en el segundo. Este proceso sólo puede "ver" con los dos ojos abiertos a la vez.

¿Cómo contribuye a la visión un proceso binocular puro? Para percibir profundidad tridimensional, el sistema visual busca partes comunes y ligeras diferencias entre detalles de las imágenes de cada ojo. El proceso puramente binocular puede contribuir a esta búsqueda. Sin embargo, la percepción visual no puede basarse, de una forma exclusiva, en la señal de salida de un proceso escuetamente binocular. Si así fuera, todas las partes desiguales de las dos imágenes retinianas desaparecerían. De hecho, lo que no podemos ver mediante un proceso exclusivamente binocular lo vemos a través de un proceso monocular o mediante un proceso que responde a cualquiera de los ojos.

Desde antiguo se ha hablado de cinco sentidos. No cabe ya duda alguna de que la cifra es mayor. Sentidos como el tacto parecen estar divididos en múltiples submodalidades; el sistema visual es también divisible. Quizá no deba considerarse la visión un solo sentido, sino un conjunto de sistemas, cada uno con entidad propia. Probablemente esté por descubrir todavía la dotación completa de sentidos visuales humanos. Los sentidos recuerdan los viejos desvanes: siempre se descubre algo inesperado en cualquier rincón que hurguemos.

La corona solar

Visible sólo durante los eclipses, la perlada envoltura del Sol se halla sometida a constante vigilancia. Las observaciones revelan una intensa actividad que refleja la interacción entre la materia y los campos magnéticos

Richard Wolfson

Uno de los espectáculos de la naturaleza más raros y llamativos es la corona solar, esa delgada atmósfera externa del Sol que se hace visible durante un eclipse total. Aunque un fenómeno así viene a ocurrir cada año en algún punto de la Tierra, la mayoría de la gente sólo tiene oportunidad de presenciar uno o dos en su vida. La luz de la corona solar queda enmascarada por la luz de la superficie luminosa del Sol, o fotosfera. La corona se hace visible merced a una curiosa coincidencia: el tamaño aparente de la Luna es casi exactamente el mismo que el del Sol. Si la Luna fuese menor, nunca habría eclipses totales; si mayor, la intrincada corona interior quedaría oculta a la visión.

A lo largo de los últimos diez años hemos avanzado un gran trecho en nuestro conocimiento de la corona y del propio Sol merced al desarrollo de instrumentos instalados en tierra o a bordo de vehículos espaciales, tripulados y no tripulados. ¿Qué vemos? Una corona en flujo continuo, cortado por periódicos brotes explosivos. Un "viento" de partículas coronales mana hacia el exterior a través del sistema solar; las fluctuaciones del viento reflejan la actividad de la corona. Esta interpretación de la corona y sus manifestaciones transforma nuestro cuadro tradicional: aquel objeto compacto que ocupaba el centro del sistema solar se convierte en una entidad de dimensiones gigantescas, que trasciende los planetas más alejados. La corona constituye un laboratorio, cercano e inmenso, para el estudio de los procesos astrofísicos. Su comprensión nos ayudará a descifrar fenómenos tan dispares como nuestros reactores de fusión y las explosiones de supernova en las estrellas distantes.

En las fotografías ordinarias, la corona aparece como un halo perlado que rodea al eclipsado Sol. La mayor parte de la luz procedente del halo no se origina en la corona: es luz solar difundida por los electrones coronales. Por cada millón de fotones que abandonan la fotosfera solar, sólo uno es difundido por la corona. Esa es la explicación de que la corona sea un millón de veces más débil que la fotosfera. La transparencia de la corona indica que el gas coronal está extremadamente disperso; con el criterio de nuestros laboratorios terrestres, constituiría un excelente vacío; la máxima densidad del gas coronal está entre 10^8 y 10^9 partículas por centímetro cúbico.

Los estudios espectroscópicos sugieren que la composición de la corona es similar a la del Sol, dominada por hidrógeno y helio con sólo indicios de elementos más pesados; ofrecen también indicaciones sobre la temperatura de la corona. Cuando los gases se calientan, se ionizan sus átomos, esto es, pierden uno o más electrones de su corteza. Comparado con el gas de la fotosfera, cuya temperatura es de unos 6000 grados Kelvin (grados Celsius por encima del cero absoluto), el coronal aparece mucho más ionizado. En la corona, el hidrógeno y el helio están esencialmente despojados de electrones; los elementos más pesados han perdido muchos de ellos. El alto estado de ionización significa que en la corona reina una temperatura muy elevada. Además, las líneas del espectro coronal están ensanchadas, lo que indica que las partículas se encuentran allí en movimiento desordenado a grandes velocidades. Fundados en esas observaciones espectroscópicas, se estima que la temperatura de la corona se sitúa entre uno y dos millones de grados Kelvin.

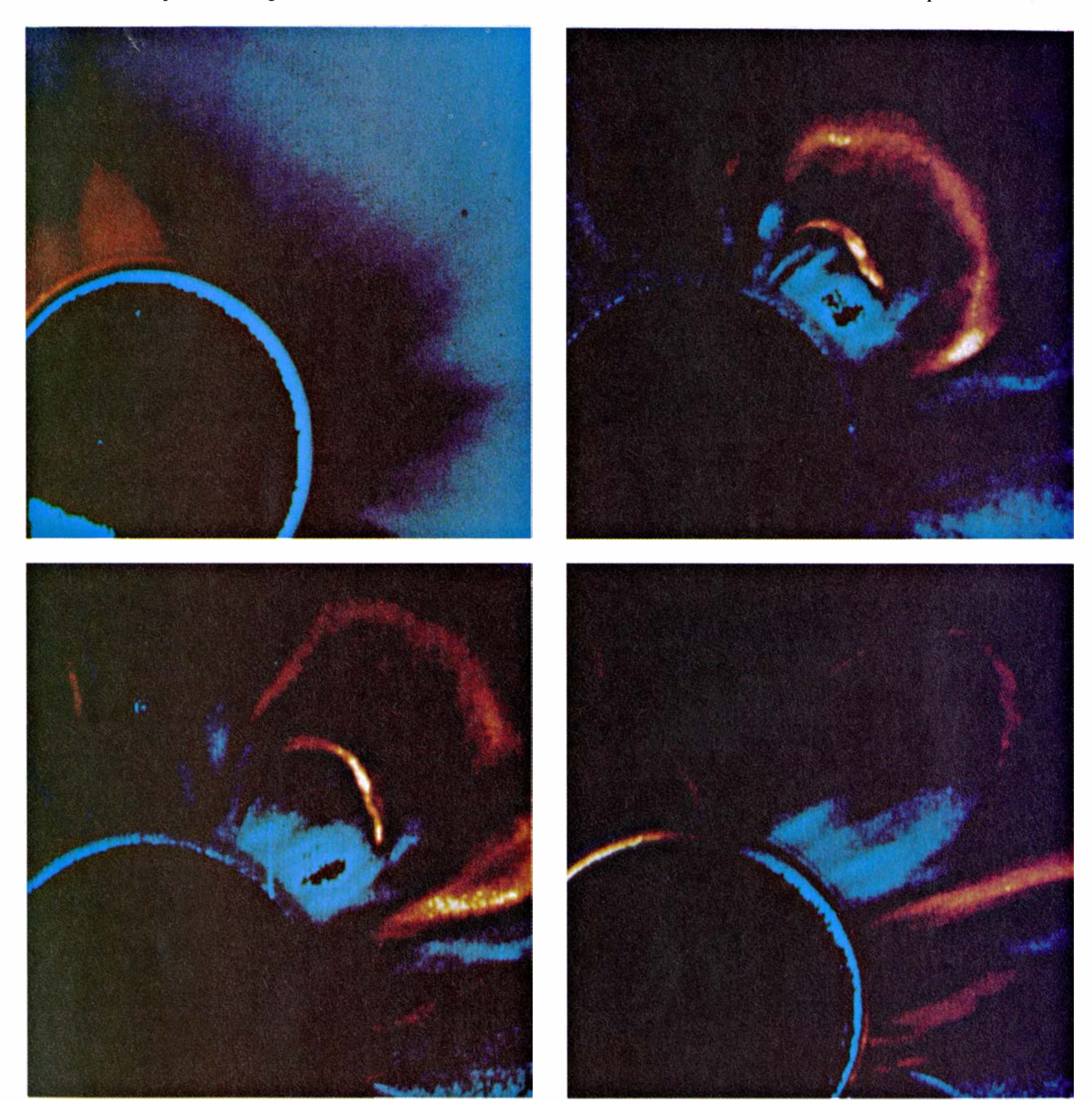
Durante largo tiempo, explicar la alta temperatura de la corona constituyó un auténtico rompecabezas para los teóricos. De acuerdo con el segundo principio de la termodinámica, el calor no puede pasar espontáneamente de un cuerpo frío a otro más caliente; aplicado a nuestro caso, significa que la energía térmica del Sol, a pesar de su enorme magnitud, no podrá calentar la corona más allá de 6000 grados Kelvin. Ante tamaña y sólida objeción, transcurrieron decenios en que los físicos se resistieron a aceptar sin más las indicaciones espectroscópicas sobre la temperatura. Se avanzaron diversas hipótesis al respecto; se sugirió, por ejemplo, que los iones de la corona podrían originarse en la profundidad interna del Sol, o acaso producirse directamente en la corona por fisión nuclear. En el decenio de 1940 la mayoría aceptaba ya la idea de una corona caliente. Una vez que los teóricos se convencieron de que allí reinaban altas temperaturas, se concatenaron los adelantos que darían cuenta de los grandes rasgos de la estructura física de la corona.

Si el segundo principio de la termodinámica prohibe el intercambio de energía calorífica desde la fotosfera hasta la corona, ¿cuál es el mecanismo de calentamiento? La respuesta más probable está en el intercambio directo de energía mecánica. La energía mecánica que se origina en la capa convectiva del Sol bajo la fotosfera es transportada hasta la corona en forma de intensas ondas sonoras o de ondas magnetohidrodinámicas.

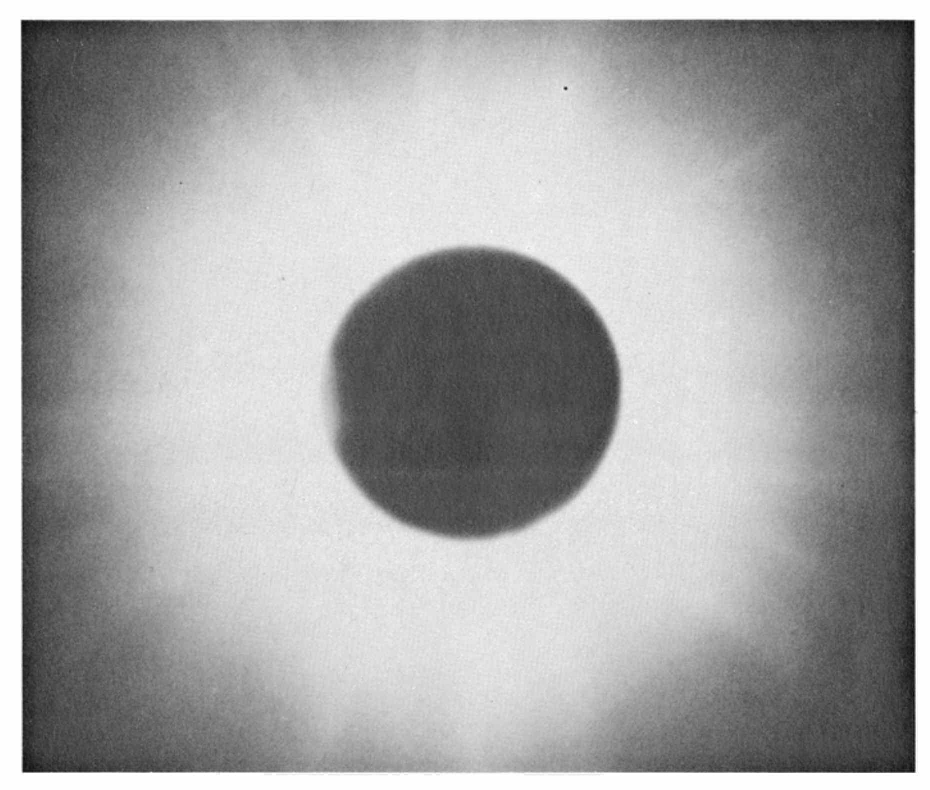
Las ondas sonoras son simples perturbaciones mecánicas de un gas que se manifiestan en forma de variaciones de presión. Las ondas magnetohidrodinámicas, que se producen en gases ionizados, son perturbaciones producidas por la interacción entre materia ionizada y campos magnéticos. La idea general de calentamiento mecánico por las ondas está sólidamente asentada; pero no se dispone de un modelo indiscutido y concordante con las observaciones. En

su diseño se afanan hoy muchos investigadores. El trabajo promete enlazar las propiedades físicas de la corona con las de las capas interiores. Se trata, en definitiva, de conseguir un marco teórico unificado del Sol en su integridad.

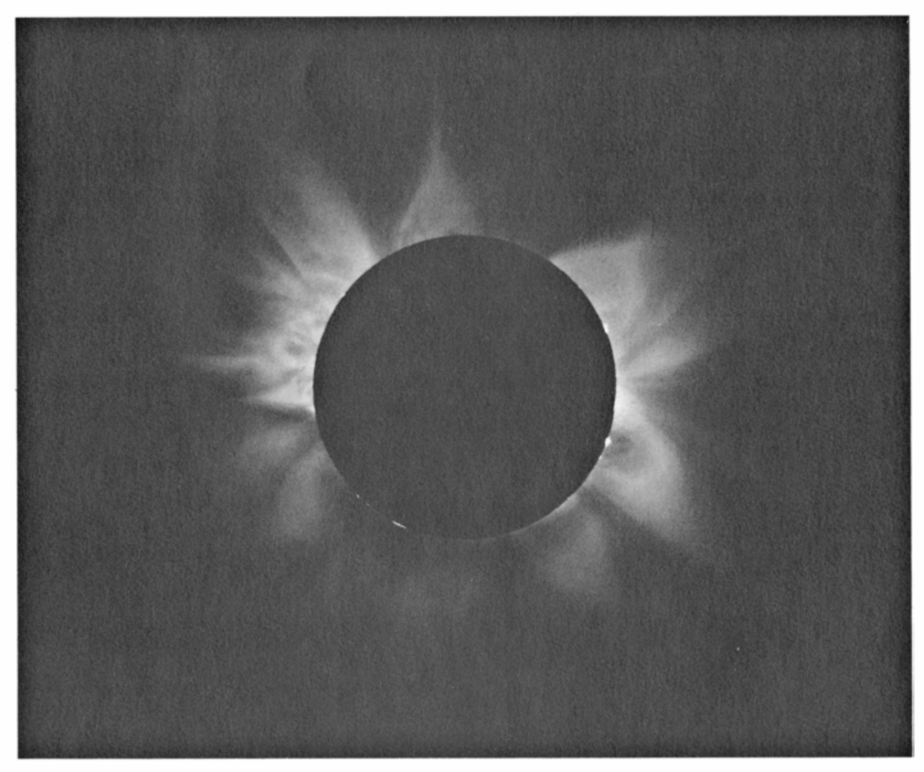
La corona es, pues, una región de alta temperatura y baja densidad que rodea a la fotosfera, más fría, de densidad mayor y mejor conocida por todos. (Hay una delgada capa intermedia, la cromosfera, que separa la fostofera de la corona.) La mayoría de las fotografías muestran una corona asimétrica, debilitada y exigua en la vecindad de los polos solares. Lejos de los polos, prominentes abultamientos sugieren una estructura muy compleja. Tales desviaciones de la simetría varían de un eclipse a otro. Las coronas más simétricas coinciden con las épocas de máxima



1. BUCLE TRANSITORIO EN LA CORONA SOLAR registrado en abril de 1980 por el coronágrafo a bordo del satélite de la Misión del Máximo Solar (SMM). Un coronágrafo lleva un disco que impide el paso de la brillante luz procedente de la fotosfera solar, o superficie visible, de manera que se pueda observar la corona, un millón de veces más débil. El coronágrafo que transportaba el vehículo espacial SMM se diseñó en el Observatorio de Gran Altitud del Centro Nacional de Investigación Atmosférica en Boulder, Colorado. La luz de la corona se focaliza sobre el tubo de una cámara de televisión; la imagen se transmite a la Tierra en forma digital. En esta secuencia en cuatro partes, presentada en colores arbitrarios, la corona antes del suceso transitorio aparece en la parte superior izquierda. Las otras tres imágenes se han construido mediante una técnica de sustracción digital, en la cual el valor de la luz en cada punto representa la variación de la intensidad luminosa con respecto al valor registrado en la primera imagen. En la parte superior derecha, el bucle coronal ha alcanzado casi el diámetro del Sol. En la parte inferior izquierda, 25 minutos más tarde, el bucle, dilatándose a una velocidad de unos 300 kilómetros por segundo, ha alcanzado el borde de la figura y, detrás de él, una prominencia solar ha comenzado a moverse hacia arriba a través de la corona. En la parte inferior derecha, 20 minutos más tarde, sólo se deja ver una de las porciones inferiores del bucle. No se sabe si la erupción del bucle coronal y la prominencia guardan una relación causal o si ambas son desencadenadas por alguna inestabilidad subyacente en el campo magnético del Sol.



2. EN UNA FOTOGRAFIA CORRIENTE DE UN ECLIPSE SOLAR, la corona interna queda sobreimpresionada y se pierden todos los pormenores. La fotografía se realizó en la India el 16 de febrero de 1980 por un grupo del Williams College dirigido por Jay M. Pasachoff. Se ve aquí que la corona es simétrica, característica ésta asociada con la máxima actividad en el ciclo de manchas solares de 11 años.



3. ESTRUCTURAS DE LA CORONA, según se vieron durante el eclipse del 16 de febrero de 1980 y captadas interponiendo un filtro, graduado radialmente, entre la lente del telescopio y la placa fotográfica. Se recogen, pues, los detalles cegados en la foto anterior. Reduciendo la brillante luz procedente de la corona interna, el filtro permite registrar las características coronales desde la fotosfera hasta una distancia de varios radios solares de ésta. Los puntos brillantes que cortan el disco lunar son prominencias de materia fría y densa soportada en la corona por los campos magnéticos. Pueden verse muchas serpentinas coronales. La fotografía fue tomada en la India por un grupo del Observatorio de Gran Altitud.

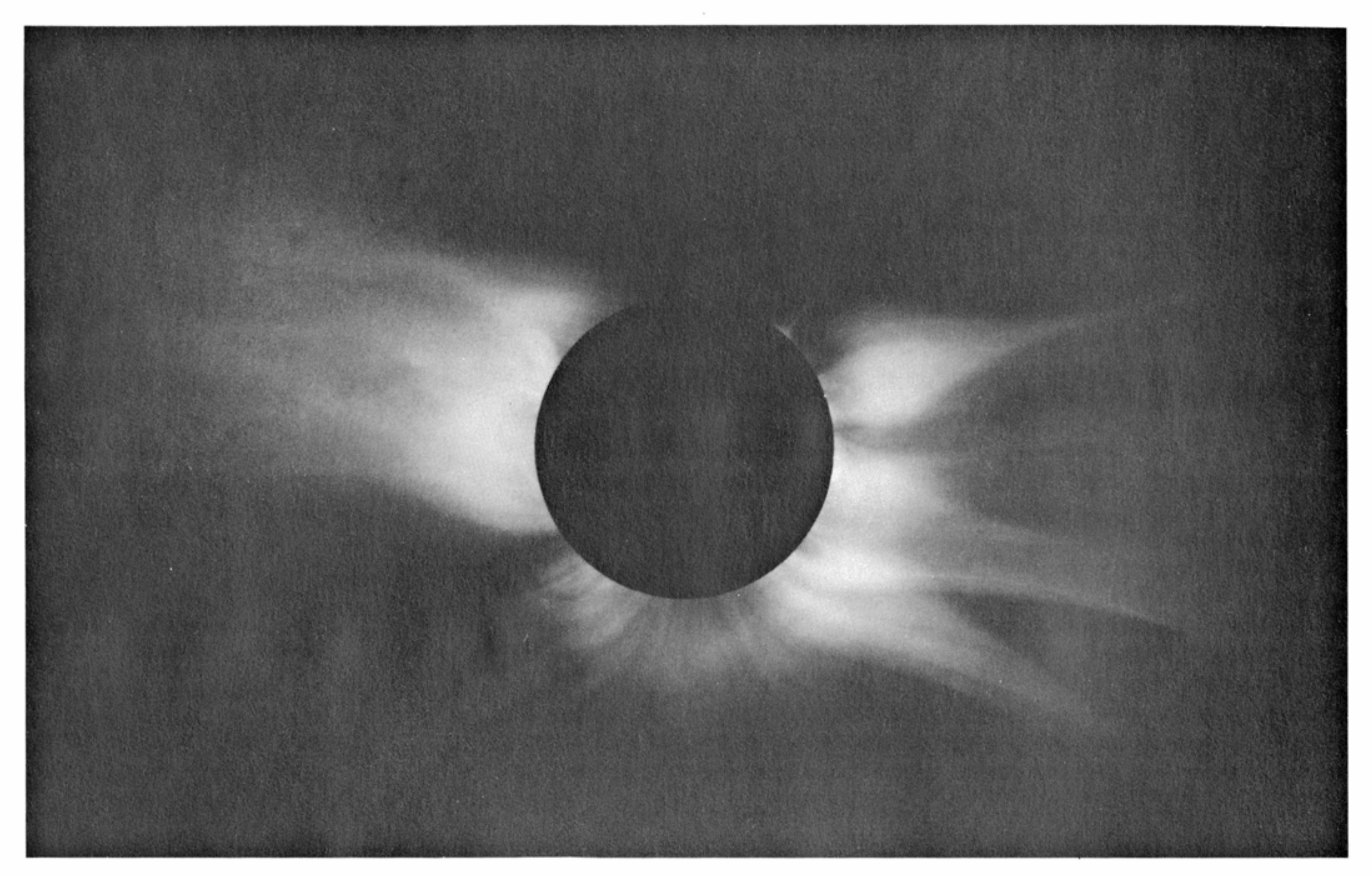
actividad de manchas solares en el ciclo solar de 11 años.

La estructura coronal insinuada en las fotografías ordinarias queda en gran parte borrada por sobreexposición de la corona interior, más brillante. El método idóneo para superar esta limitación consiste en fotografiar la corona por medio de filtros especiales de densidad graduada radialmente, de suerte que absorban más luz en el centro de la imagen que hacia el borde. De este modo se compensa la disminución del brillo de la corona con la distancia hacia fuera. Tales filtros permiten detectar las estructuras coronales a muchos radios solares de distancia desde la fotosfera.

En una fotografía tomada mediante un filtro graduado radialmente, la delicada belleza de la corona se nos ofrece en curiosos pormenores. Largas y graciosas serpentinas se extienden millones de kilómetros hacia fuera. Cerca del disco brillan lazos de materia coronal. En algunas fotografías se contemplan también rayos estrechos y delgados. En las épocas de máxima actividad solar, esas figuras rodean todo el disco solar. En épocas de mínima actividad, las características coronales quedan, por el contrario, confinadas a las latitudes solares más bajas. La corona muestra entonces grandes "agujeros" en ambos polos, donde la corona es mucho más débil que en la proximidad del ecuador. Dentro de los agujeros polares se distinguen frecuentemente débiles "penachos" polares. Desde cada polo arranca un conjunto de esas estructuras delgadas que van debilitándose hasta perderse a un radio solar de distancia, aproximadamente. Las figuras dominantes –serpentinas, lazos y agujeros– ocupan de tal manera el panorama que algunos investigadores creen que la corona consta sólo de esas estructuras y no existe nada que hiciera sospechar una "corona tranquila" subyacente.

El aspecto de la corona varía bastante de un eclipse a otro. Las fotografías de eclipses difieren tanto de un caso a otro que los observadores identifican de qué episodio se trata con sólo la fotografía.

Las imágenes coronales tomadas mediante filtros graduados radialmente varían de una forma espectacular. Además de los cambios generales asociados al ciclo de las manchas, cada eclipse tiene su propia configuración de serpentinas, lazos y otras características. Un conjunto de fotografías hechas en diferentes momentos a lo largo de un mis-



4. CORONA ASIMETRICA, fotografiada mediante un filtro graduado radialmente durante el eclipse del 30 de junio de 1973, cuando el ciclo de 11 años de las manchas solares estaba cerca de un mínimo. En tales épocas, las serpentinas coronales se concentran cerca del ecuador solar y los polos muestran "agujeros", o regiones de actividad mínima. La fotografía fue tomada en Kenya por un grupo de astrónomos adscrito al Observatorio de Gran Altitud.

mo eclipse presenta una preocupante indicación de una corona dinámica, cuyo aspecto cambia en una escala de tiempo mucho más corta que el intervalo entre dos eclipses.

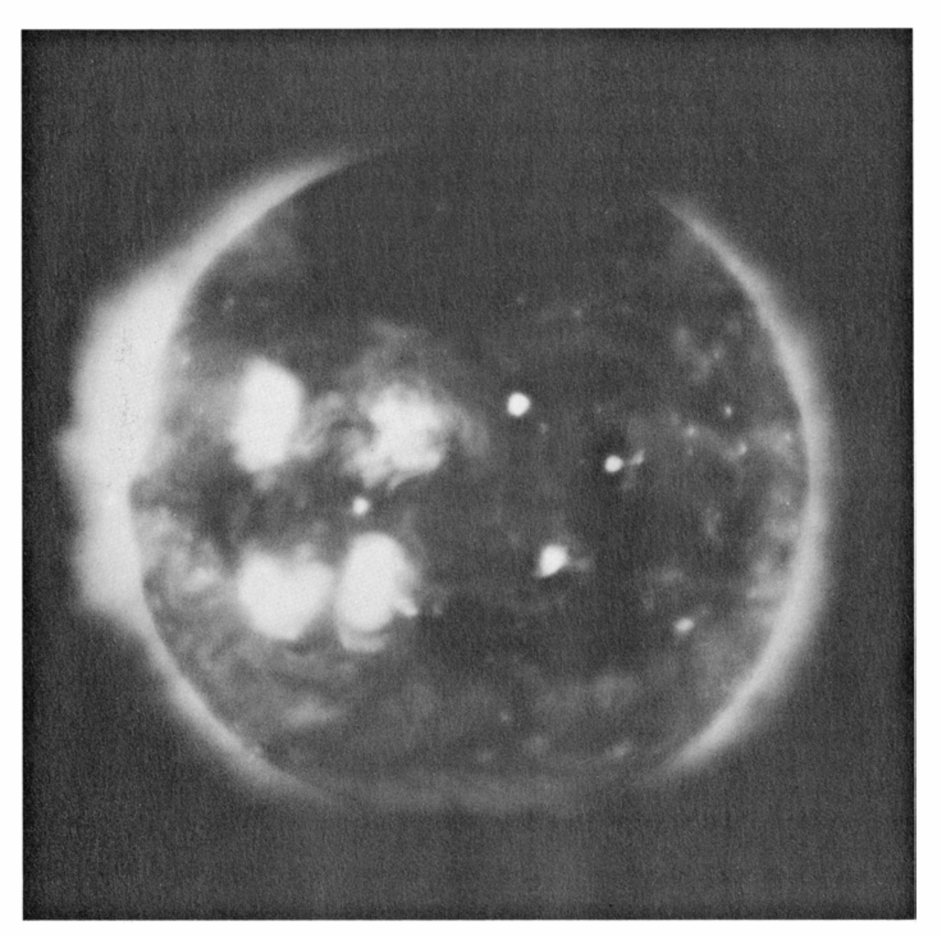
Para estudiar la dinámica de la corona, sería deseable observarla también cuando no hay eclipse. Considerado el brillo de la fotosfera solar, ello parece imposible. A las longitudes de onda del visible, la densa fotosfera brilla un millón de veces más que la tenue corona. Por ser ésta mucho más caliente que aquélla, produce radiación electromagnética de alta energía que apenas si existe en el espectro de la fotosfera. A dos millones de grados, el gas coronal constituye una intensa fuente de rayos X. Si descontamos los rayos X emitidos esporádicamente por las fulguraciones solares y otras "manchas calientes" locales, la fotosfera, más fría, es una fuente de rayos X de importancia despreciable. Podemos observar, pues, la corona dirigiendo al Sol un telescopio de rayos X.

Ese tipo de observaciones se realizan desde el espacio, ya que la atmósfera terrestre es opaca a los rayos X. Los primeros telescopios de rayos X se elevaron en cortos vuelos con cohetes, a principios del decenio de 1970; tomaron unas imágenes toscas de la corona que mostraban áreas brillantes, lo que sugería la existencia de regiones de alta densidad o temperatura. Comparando las fotografías tomadas desde tierra durante eclipses con las imágenes de rayos X hechas desde el espacio en aproximadamente el mismo instante, se confirmó la correspondencia entre las regiones brillantes en rayos X y ciertos rasgos de la corona que se captaban con medios ópticos.

La técnica de la construcción de imágenes de rayos X ha experimentado un rápido progreso a lo largo de los últimos diez años. Hacia mediados de la década de 1970, los astronautas a bordo del observatorio orbital Skylab obtuvieron muchas imágenes de rayos X de la corona. Una película confeccionada a partir de tales imágenes nos revela con claridad cómo evolucionan las figuras coronales a medida que procede la rotación solar. Las imágenes de rayos X del Skylab resultaron decisivas para establecer la existencia de agujeros coronales, no sólo en los polos, sino también, aunque en menor escala, en toda la corona. En vehículos espaciales más recientes, métodos mejorados de formación de imágenes de rayos X e instrumentos espectroscópicos han suministrado datos de la corona de rayos X de resolución todavía superior.

A partir de la radiación ultravioleta emitida por elementos pesados y altamente ionizados de la corona, podemos extraer información adicional de la naturaleza de ésta. A tales longitudes de onda es también despreciable la emisión de la fotosfera. El valor de los datos ultravioleta reside en que las líneas espectrales particulares se forman sólo dentro de estrechos márgenes de temperatura, lo que nos posibilita explorar diferentes capas de la corona, la cromosfera y la región de transición entre ambas.

En el otro extremo del espectro, los radiotelescopios basados en tierra pueden identificar y seguir señales asociadas a oscilaciones características del gas coronal ionizado. La frecuencia de tales oscilaciones constituye una medida directa de la densidad de la corona: se pueden explorar regiones coronales diferentes con solo cambiar la frecuencia



5. IMAGEN DEL SOL EN RAYOS X, que registra radiación originada predominantemente en la corona, donde la temperatura es de unos dos millones de grados Kelvin (grados Celsius por encima del cero absoluto). La fotosfera solar, a 6000 grados Kelvin, resulta demasiado fría para emitir rayos X. La imagen de rayos X fue tomada por los astronautas a bordo del Skylab el 18 de septiembre de 1973, dos meses y medio después del eclipse solar presentado en la figura 4. La mayoría de las regiones brillantes de la imagen de rayos X son zonas donde se incrementan la densidad o temperatura coronales, correspondiendo a serpentinas coronales o a bucles. Las zonas oscuras son agujeros coronales. Además de los agujeros polares, se distingue un agujero coronal prominente cerca del ecuador solar. Imágenes así demuestran que los agujeros coronales no se ciñen a las regiones polares del Sol, sino que están asociados también a chorros de gran velocidad en el viento de partículas solares. La imagen nos la cedió la American Science & Engineering, Inc., que proyectó el telescopio de rayos X, instalado a bordo del Skylab.

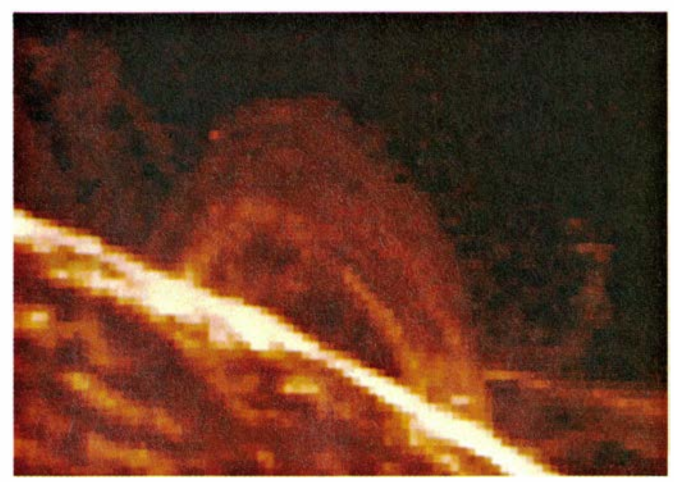
de sintonía del radiotelescopio. Uniendo varios radiotelescopios para constituir un interferómetro, podemos seguir la evolución de las distintas figuras. En Culgoora, Australia, la Organización de Investigación Científica e Industrial (CSIRO) ha montado un radioheliógrafo, que consta de 96 radiotelescopios interconectados y dedicados a estudios solares.

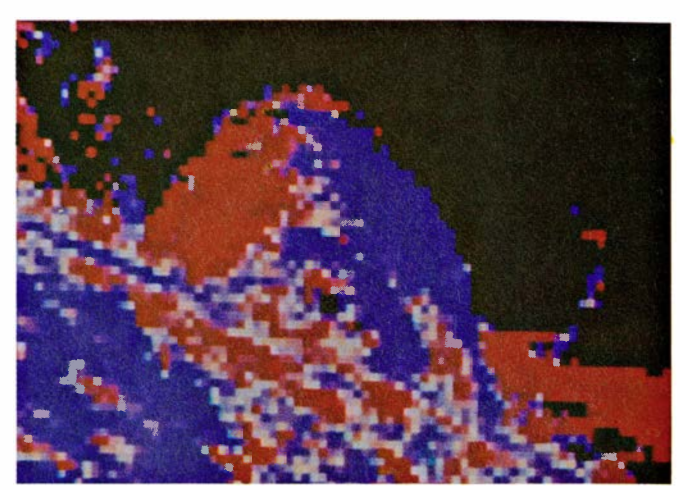
El método más directo para investigar la corona en ausencia de eclipse es simular un eclipse artificial. Bastaría en principio con mantener el pulgar derecho del brazo estirado para ocultar la luz de la fotosfera. El método no da resultado porque la atmósfera terrestre difunde suficiente luz del Sol para enmascarar la luz coronal. Por encima de la atmósfera, sin embargo, e incluso en la cima de montañas altas, se observa la corona impidiendo artificialmente el paso de la luz de la fotosfera. Se ha ideado un instrumento para ese fin, el coronágrafo. Consta de un pequeño telescopio refractor equipado con un disco que oculta el Sol, de la misma manera que la Luna lo hace en un eclipse; lleva incorporados varios dispositivos de ocultación proyectados para reducir al máximo la luz difusa y los efectos de difracción que distorsionan la imagen coronal. Una vez aislada, se registra directamente la luz coronal, o se la hace pasar a través de polarizadores o filtros para determinar características específicas.

Los primeros coronágrafos se desarrollaron hace 50 años. Desde entonces se han instalado en cimas de montañas, han volado en globos y se han lanzado al espacio. Un coronágrafo del Skylab descubrió los coronales transitorios, o estructuras rápidamente cambiantes que parecen expulsar material coronal hacia el espacio. Hay coronágrafos volando a bordo de varios vehículos espaciales no tripulados. De los instalados en tierra, los más importantes están en el Observatorio del Pico Sacramento en Nuevo México, en el Pic du Midi en Francia y en Mauna Loa en las islas Hawai. En 1979, un coronágrafo orbital asignado al Laboratorio de Investigación Naval de los Estados Unidos registró un suceso nunca observado hasta entonces: el choque de un cometa contra el Sol. Los más modernos registran sus imágenes mediante detectores electrónicos digitales y no en película fotográfica. Los datos de los detectores electrónicos se pueden tratar en ordenador para dar imágenes realzadas con alta resolución y contraste.

Algunas de las mejores imágenes coronales provienen de un coronágrafo montado en el satélite de la Misión del Máximo Solar (SMM), lanzado en febrero de 1980, cerca del máximo de actividad solar dentro del ciclo actual de manchas solares. El vehículo espacial lleva un conjunto de instrumentos proyectados para realizar un estudio coordinado del Sol a distintas longitudes de onda: desde el visible hasta las de rayos gamma. Durante algo más de seis meses, el SMM aportó un verdadero caudal de datos, hasta que se produjo cierto número de fallos en el sistema que dirige el vehículo espacial hacia el Sol y éste quedó girando sin control; así se abortó una misión prevista para que durase dos años. Pero ocurre que el SMM es el primer vehículo espacial susceptible de ser reparado en el espacio. Está equipado con un garfio de enganche y sus paquetes de instrumentos son intercambiables. La misión Transbordador Espacial, prevista para abril de 1984, incluirá un encuentro con el SMM durante el cual un astronauta se acercará al satélite averiado y detendrá su giro. En cuanto quede estabilizado el satélite, el brazo mecánico del transbordador lo recogerá y lo situará en su bodega de carga donde se reparará para ponerlo en órbita de nuevo. El SMM reemprenderá su misión durante el mínimo de manchas solares, lo que habrá de aportar valiosos datos a comparar con los recogidos durante el máximo.

El desarrollo del coronágrafo del SMM corrió a cargo de un grupo dirigido por Lewis L. House, del Observatorio de Gran Altitud del Centro Nacional de Investigación Atmosférica, Colorado. En él, se hace pasar la luz coronal a través de filtros intercambiables y polarizadores y se focaliza sobre el tubo de una cámara de televisión.

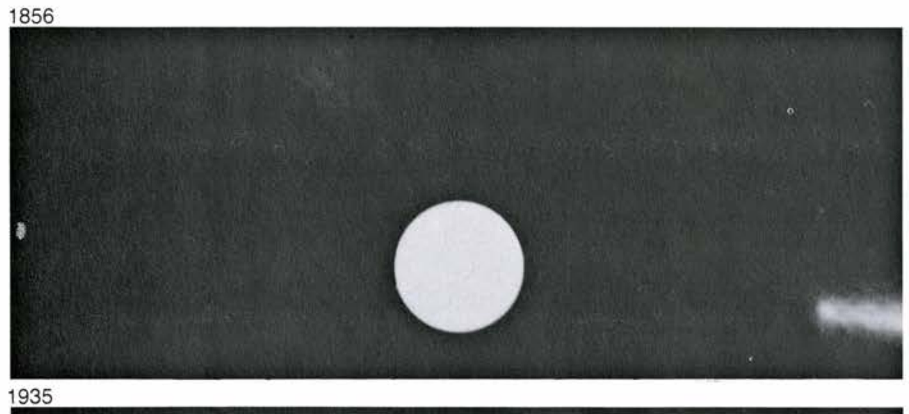




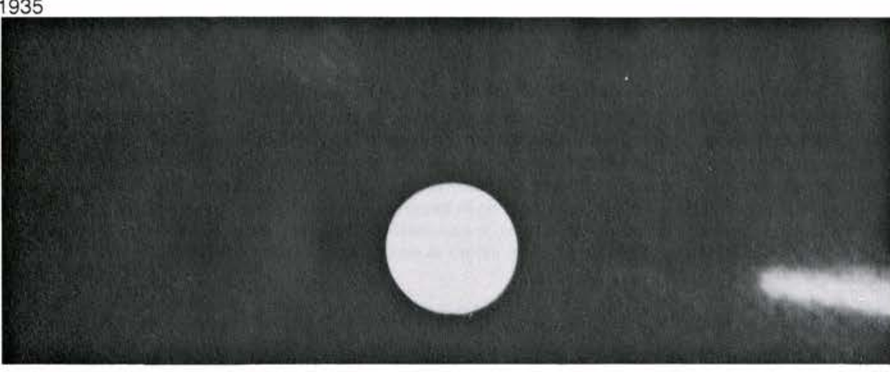
6. BUCLE CERCA DE LA FOTOSFERA, soportado por un intenso campo magnético local. Registra una temperatura próxima a los 100.000 grados Kelvin. Las imágenes recogen la radiación ultravioleta a una longitud de onda de 1548,19 angstrom emitida por el carbono triplemente ionizado, esto es, átomos de carbono despojados de tres electrones de los seis que normalmente tienen. La imagen de la izquierda está tomada con toda la radiación de una estrecha banda centrada a esa longitud de onda. La imagen de la derecha es un "diagrama de velocidad" del mismo bucle, formada separando la radiación en dos longitudes de onda componentes, una ligeramente más corta y otra un tanto mayor que la longitud de onda central. La componente más corta, en azul, proviene de materia que se está acercando al instrumento observador (espectrofotómetro ultravioleta de la MMS); la componente de mayor longitud de onda, en rojo, proviene de materia que se está alejando. Una interpretación de la imagen dice que la mitad derecha del bucle es la más cercana de las dos y que existe materia en movimiento descendente hacia ambas patas. Esto revela el carácter dinámico de las capas externas del Sol.

7. SERPENTINAS CORONALES Y UN BUCLE ascendente con prominencia, registrados por el coronágrafo del Observatorio de Gran Altitud, instalado a bordo del satélite de la Misión del Máximo Solar, el 14 de abril de 1980. Se realzó el contraste de la imagen mediante "delineación efectiva", técnica de ordenador para destacar los detalles. Los distintos colores de la imagen corresponden a diferentes valores de la densidad de la corona. La imagen es una versión de alto contraste del suceso coronal que aparece en la secuencia en cuatro partes reproducida en la primera ilustración de este artículo.

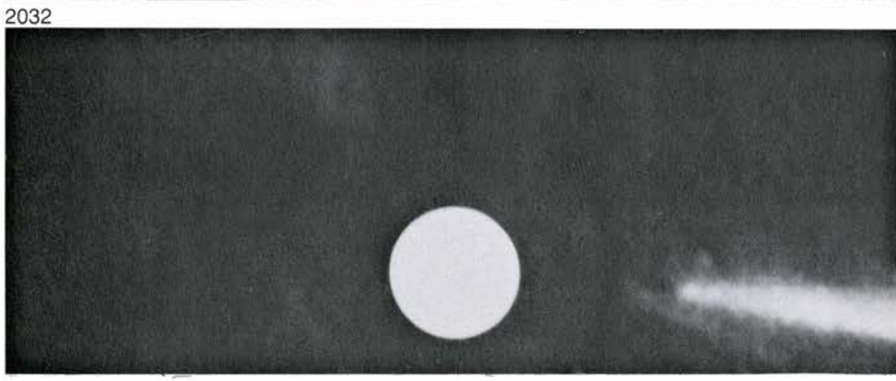
1856



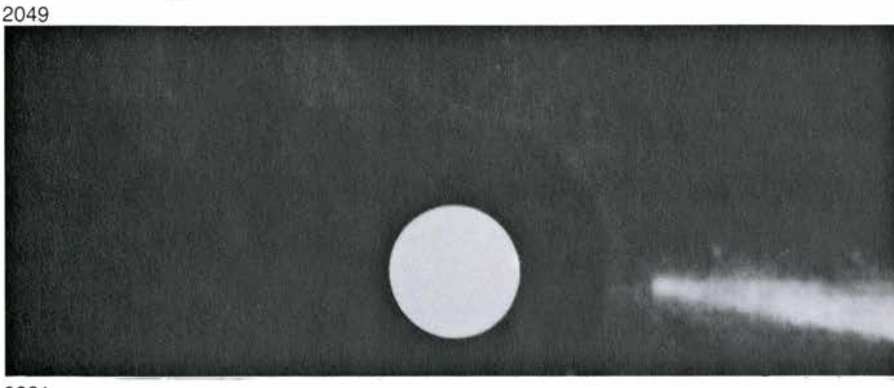
1935



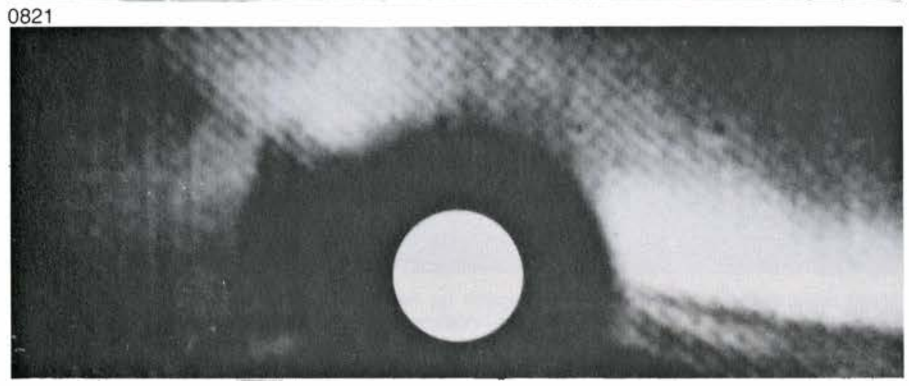
2032



2049



0821



8. CHOQUE DE UN COMETA con el sol, primer suceso de su clase registrado. Lo fotografió un coronágrafo orbital dependiente del Laboratorio de Investigación Naval de los Estados Unidos. La secuencia comienza (*arriba*) a las 1856, Tiempo Universal, el 30 de agosto de 1979. El disco del coronágrafo tapa al Sol; el disco blanco representa la posición y tamaño del Sol. El cometa, uno de la familia conocida como "rasantes al Sol", nunca fue detectado por los observadores. El resplandor después del "chapuzón" persistió durante 24 horas. Desde 1979 se han observado dos choques más.

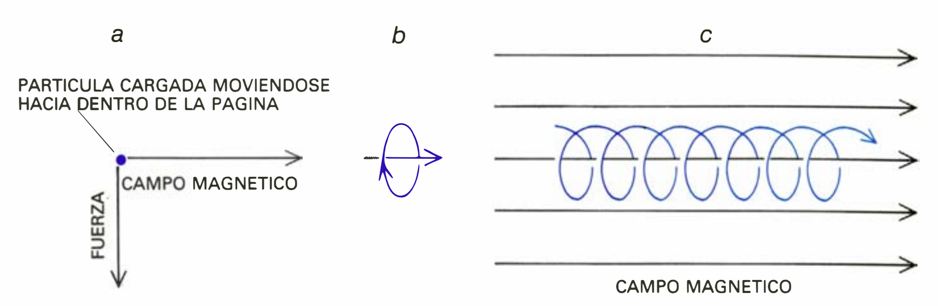
Los datos de la cámara se transmiten a tierra en forma digital y se almacenan en cinta magnética. Cada imagen supera el millón de bits de información y recoge un cuadrante de la corona. Se procesan las imágenes en un ordenador de tierra y se proyectan en un monitor de televisión en color. El observador se sirve del ordenador para ampliar regiones seleccionadas de la imagen y realzar las estructuras que le interesen.

Antes de su interrupción temporal en el otoño de 1981, el coronágrafo del SMM envió unas 30.000 imágenes coronales. En una película, preparada por Rainer M. E. Illing, del Observatorio de Gran Altitud, a partir de una secuencia de imágenes tomadas en abril de 1980, se nos revela una corona "nueva": cual una estructura, inquieta y dinámica, en perpetuo cambio de aspecto. Aunque los rasgos coronales más importantes persisten y pueden reconocerse durante varios días o más, salta a la vista que evolucionan de un día para otro. Las figuras más pequeñas surgen y desaparecen en un lapso más breve todavía. Episodios transitorios ocasionales deforman la corona en cuestión de horas. Un suceso fugaz típico, captado por el coronágrafo SMM, nos muestra un lazo coronal que se ensancha de golpe e incurva hacia fuera a una velocidad de cientos de kilómetros por segundo.

¿Cuál es la naturaleza física de las figuras coronales? ¿Cómo puede un gas caliente y tenue organizarse en configuraciones tan bien definidas? ¿Por qué algunos rasgos persisten durante días, en tanto que otros evolucionan brusca y rápidamente? Ni la fuerza de la gravedad, esféricamente simétrica, ni las fuerzas asociadas a la rotación del Sol pueden moldear la corona en sus intrincadas formas. Las propias figuras dan una pista sobre el origen de las estructuras coronales. Los lazos y los penachos polares recuerdan las formaciones que se producen cuando se espolvorean limaduras de hierro en la proximidad de un imán. De igual ma-

nera que las limaduras se alinean a lo largo de las líneas de fuerza magnética, la estructura del campo magnético solar se impone sobre el gas coronal. La fuerza magnética es la responsable del detalle, variedad y evolución de las estructuras coronales.

Como el campo magnético de la Tierra y el de otros muchos objetos celestes, el solar nace de la interacción entre el movimiento convectivo y la rotación del material eléctricamente conductor del interior del objeto. Por razones que todavía no se conocen bien, el campo magnético solar es muy variable. Aproximadamente cada 11 años, los polos magnéticos Norte y Sur del Sol se permutan. Los polos magnéticos presentan su máxima intensidad en la vecindad del mínimo de manchas solares y confieren al campo magnético solar la estructura general de un dipolo magnético, cual si de una barra imantada o de la propia Tierra se tratara. La estructura dipolar se torna evidente en los penachos polares que aparecen en la corona cerca del mínimo de manchas solares. En la vecindad del máximo de manchas solares, el campo magnético solar es caótico y no presenta polos definidos, sino más bien muchas regiones locales de intenso magnetismo, asociadas frecuentemente a las manchas solares. Los bucles coronales, que en esta época proliferan, definen regiones de intensos campos magnéticos locales.



9. PARTICULAS CARGADAS de la corona. Presentan el comportamiento característico de todas las partículas cargadas bajo la influencia de un campo magnético. La fuerza que actúa sobre una de ellas es perpendicular a la dirección del campo y a la dirección del movimiento de la partícula. Para una partícula cargada positivamente que se mueva perpendicularmente hacia esta página (*a*) en presencia de un campo magnético dirigido hacia la derecha, la fuerza se encaminaría hacia abajo. A medida que la partícula se moviera, la fuerza la desviaría continuamente de la trayectoria rectilínea, de modo que acabaría describiendo una órbita circular (*b*). Si la partícula tuviera una componente de velocidad en la dirección del campo, su trayectoria se convertiría en una hélice (*c*). La partícula se puede mover en cualquier sentido paralelo al campo; pero en respuesta a una fuerza que actúa perpendicularmente al campo, se traslada a una hélice de mayor diámetro. Una partícula de carga negativa giraría en sentido opuesto a la dibujada.

La interacción entre el gas coronal y el campo magnético se rige por las leyes del electromagnetismo. De los fenómenos magnéticos el más familiar es el comportamiento de los metales ferrosos, las limaduras por ejemplo. Pero si atendemos a su naturaleza fundamental, habrá que tener presente que el campo magnético se origina por el movimiento de cargas eléctricas. En materiales magnéticos, las cargas son electrones que se mueven en órbitas alrededor de los núcleos atómicos, produciendo colectivamente el magnetismo del trozo de materia de que se trate. En un gas ionizado, así la corona, las partículas que constituyen el gas están cargadas, por cuya razón interaccionan con campos magnéticos. Interacciones electromagnéticas que hacen de los gases ionizados algo muy distinto de los gases ordinarios; hasta el punto de que un gas ionizado ha adquirido un nombre característico: plasma. El estado de plasma se llama cuarto estado de la materia. En la Tierra, planeta relativamente frío, los plasmas escasean. Pero en otras regiones del universo la mayor parte de la materia existen en forma de plasma, según se cree.

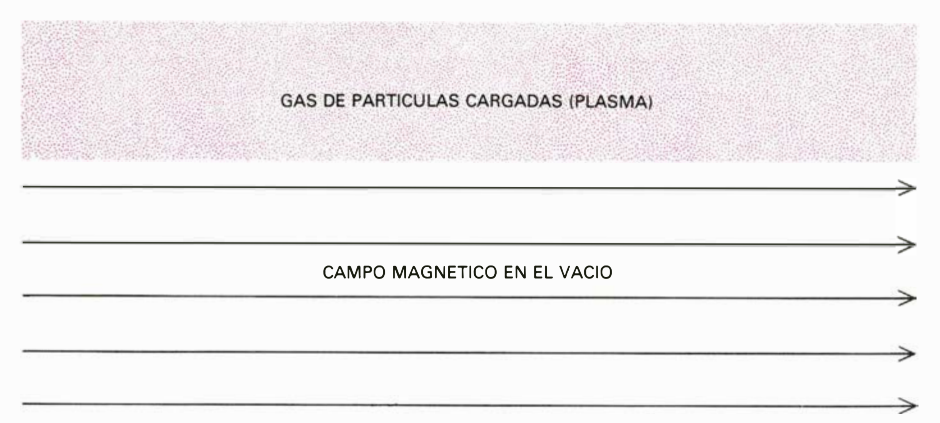
Una partícula cargada que se mueva en el seno de un campo magnético queda sometida a una fuerza que depende de su carga, de su velocidad y de la intensidad y dirección del campo. La fuerza es máxima cuando las direcciones del movimiento y del campo forman ángulo recto; disminuye hasta anularse, cuando las direcciones del movimiento y del campo corren paralelas. Por ser la fuerza siempre perpendicular a la dirección del movimiento de la partícula, sólo podrá modificar dicha dirección y no la velocidad de la partícula. Y como no hay fuerza cuando el movimiento es paralelo al campo, las únicas componentes del movimiento de la partícula afectadas serán las perpendiculares al campo.

Resulta en definitiva que las partículas cargadas se mueven describiendo trayectorias helicoidales alrededor de las líneas de fuerza del campo magnético. Las partículas se mueven libremente en la dirección del campo, pero si las empujamos perpendicularmente al mismo, describirán una circunferencia y no una línea recta. Se dice que las partículas cargadas en un campo magnético están "ligadas" al campo, de manera muy similar a las cuentas en un alambre. Las partículas se mueven libremente sólo a lo largo de la dirección del campo. Los campos no uniformes y los choques hacen la ligadura imperfecta, pero en un plasma enrarecido, como el de la corona, la hipótesis de que el plasma está ligado al campo ("congelado") es un excelente punto de partida.

Las partículas cargadas de la corona constituyen un gas. Este ejerce una presión, del estilo de la que realiza la atmósfera terrestre. Imagínese el gas coronal tratando de moverse, bajo el empuje de la presión, hacia una región donde existe un campo magnético. Como el plasma no puede moverse fácilmente a través de las líneas de fuerza del campo, éste equilibra de manera eficaz la presión del gas. Un análisis más cuantitativo muestra que un campo magnético ejerce una presión sobre el gas ionizado, presión magnética que es proporcional al cuadrado de la intensidad del campo. El comportamiento del gas coronal reviste una complejidad mucho mayor que el de la atmósfera terrestre; en efecto, la dinámica del gas viene determinada por la presión del gas y la presión magnética.

El plasma coronal no se limita a responder pasivamente a la acción del campo magnético solar. Su alta ionización y ligereza posibilita una corona que sea un excelente conductor eléctrico y mantenga corrientes eléctricas intensas. Las corrientes dan lugar, a su vez, a campos magnéticos que modifican los campos originales. Esta doble interacción entre la corona y el campo magnético del Sol confiere a la estructura coronal una riqueza y variedad impensable en un gas no conductor.

La riqueza de los fenómenos coronales complica los esfuerzos por justificar teóricamente el comportamiento de la corona. Una descripción cabal de la corona requiere resolver antes un complicado sistema de ecuaciones en las cuales el campo magnético da forma a las estructuras coronales al mismo tiempo que las corrientes coronales están modificando el campo. Además, no se pueden despreciar los efectos de la gravedad, rotación y la presión gaseosa ordinaria. Ni que decir tiene que la explicación teórica completa de todas las fi-



10. SE PRODUCE PRESION MAGNETICA porque el plasma, o gas de partículas cargadas, no se puede mover fácilmente en dirección perpendicular a un campo magnético. En el sencillo caso presentado en la ilustración, existe un vacío que encierra un campo magnético junto a un plasma sometido a presión. La presión del campo magnético impide que el plasma (*arriba*) penetre en el vacío (*abajo*). Aunque se supone que la presión del plasma y la presión magnética se equilibran, la configuración no es necesariamente estable. Cualquier perturbación del campo o del plasma desencadenará una redistribución radical.

guras coronales que se ven en un eclipse particular o en una imagen de coronágrafo está todavía por llegar.

Se ha dado, ya, un primer paso importante. Nos referimos al modelo matemático desarrollado por Gerald Pneuman y Roger Kopp, del Observatorio de Gran Altitud. Muestra con nitidez el desarrollo de serpentinas coronales en el ecuador solar. Posteriores refinamientos de modelos de serpentinas ecuatoriales llevados a cabo por Pneuman, Tyan Yeh, de la Administración Nacional Oceánica y Atmosférica, y otros nos han ofrecido una explicación detallada de la estructura de las serpentinas. La investigación actual en este campo pretende, entre otros objetivos, modelar la asociación de una serpentina y un agujero coronal adyacente. En el modelo que el autor ha venido trabajando se consideraba la formación de serpentinas a altas latitudes solares.

Los descubrimientos que se han ido resgistrando a lo largo de los últimos diez años han traído un aluvión de estudios teóricos encaminados a identificar las fuerzas que subyacen a la dinámica coronal, en particular a fenómenos de la rapidez de los sucesos coronales transitorios. De acuerdo con el trabajo reciente de B. C. Low, del Observatorio de Gran Altitud, y el autor, la fuerza ascensional magnética causaría la elevación de los bucles coronales en los sucesos transitorios. De la misma manera que un trozo de madera queda sometido a una fuerza ascensional cuando se sumerge en agua, así las regiones de la corona donde el campo magnético es más intenso tienden a ser elevadas por el gas circundante. Los modelos teóricos muestran que los bucles coronales pueden tornarse inestables de repente y elevarse en el seno de la corona, en virtud del movimiento caótico del gas fotosférico de la base del bucle o en respuesta a un violento suceso fotosférico, una fulguración solar por ejemplo.

El plasma coronal y el campo magnético están estrechamente ligados entre sí. Resulta imposible mover el plasma perpendicularmente al campo magnético sin arrastrar, con él, las líneas de fuerza magnética. Significa ello que cualquier perturbación del plasma puede crear un lazo o un abultamiento en las líneas del campo. El lazo se propaga a lo largo de la línea de fuerza magnética cual si se tratara de una deformación de un muelle estirado, que va reproduciéndose por todo el muelle. Tal perturbación viajera es una onda magnetohidrodinámica, que transporta energía a medida que deja plasma y campo magnético de sus estados de equilibrio de baja energía. Como se ha indicado antes, las ondas magnetohidrodinámicas pueden ser las responsables del transporte de energía desde la capa convectiva subyacente a la fotosfera solar hasta el interior de la corona, a través de la fotosfera y cromosfera. Haciéndose cada vez más acusadas hasta convertirse en ondas de choque magnetohidrodinámicas, las ondas disipan su energía en la corona, generando así la temperatura coronal de dos millones de grados.

¿Por qué es de interés la dinámica coronal? ¿Tiene la actividad coronal alguna influencia directa en la Tierra? Si se conociera la manera de conducirse la corona, sabríamos algo del complicado comportamiento de gases calientes e ionizados en el seno de campos magnéticos. Podría entonces conocerse el comportamiento de objetos astrofísicos más distantes y extraños, entre ellos los púlsares, los quásares y las galaxias activas. Una mayor comprensión de la corona ayudaría a dominar la fusión nuclear como fuente de energía, ya que los caminos más prometedores hacia la fusión llevan consigo el confinamiento de plasma caliente por campos magnéticos.

Mayor significación encierra este otro hecho: la Tierra está inmersa en la corona. Se sabe desde finales del decenio de 1950 que la alta tempe-



11. APARECE LA FUERZA ASCENSIONAL magnética en una región de mayor intensidad del campo magnético que está interpuesta entre dos regiones de menor intensidad, cuando las tres se hallan dentro de un campo gravitatorio. La presión total en cada región es la suma de las presiones magnética y del plasma: la región de mayor intensidad del campo magnético poseerá una menor presión del plasma si ha de estar en equilibrio de presión con su entorno. Si las regiones están a la misma temperatura, menor presión implica menor densidad. Por ser la región menos densa la que menos sufre la atracción gravitacional, flota hacia arriba; esa fuerza puede explicar sucesos en los que los bucles salen hacia el espacio.

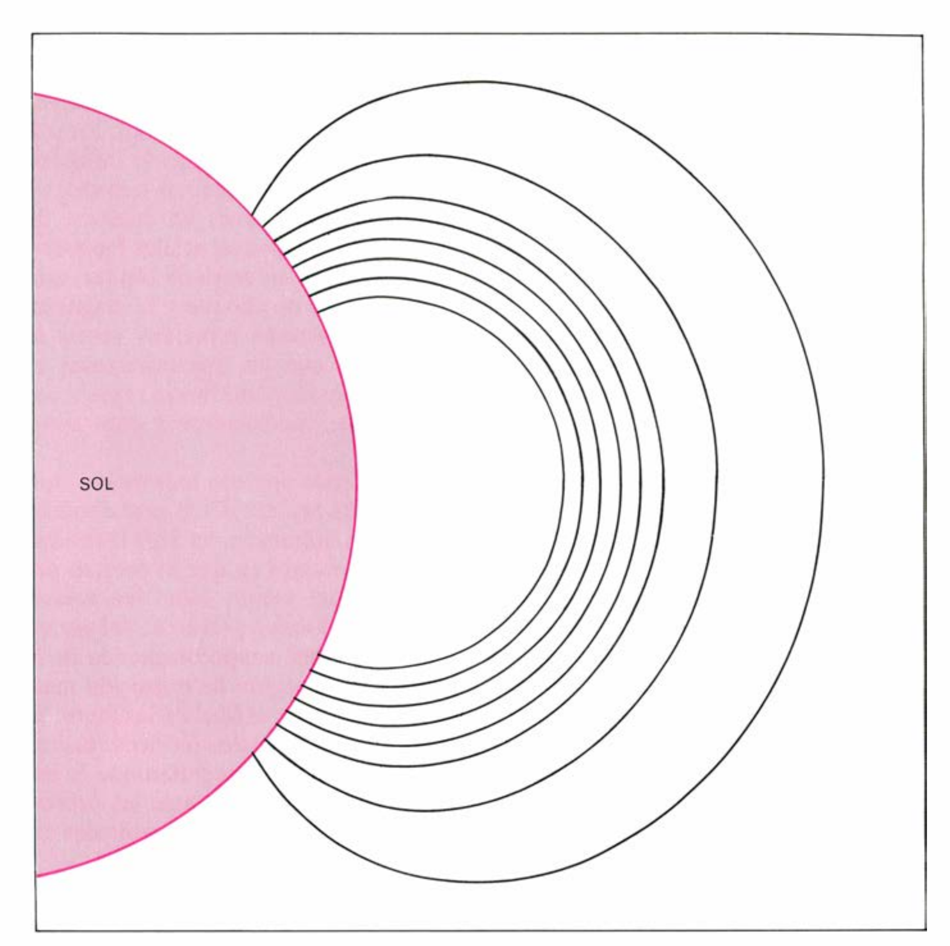
ratura de la corona produce el viento solar, chorro de material solar lanzado hacia el espacio que pasa junto a la Tierra a unos 400 kilómetros por segundo. El viento solar prolonga la corona en el espacio interplanetario.

El plasma coronal está fuertemente atrapado por el campo magnético del Sol; por tanto, aquellas partes de la corona donde se hallan inmersos los bucles magnéticos no contribuirán apenas al viento solar. Con bucles magnéticos ligados a la superficie solar en ambos extremos no hay manera de que el gas coronal del interior de los bucles escape al espacio interplanetario. En aquellas regiones donde las líneas de fuerza del campo magnético solar se extienden indefinidamente hacia fuera, sin embargo, el viento solar puede escapar a lo largo de las líneas del campo. Se cree que una porción destacada del viento solar, sus chorros de alta velocidad en particular, se origina en los agujeros coronales. Los episodios coronales transitorios pueden ofrecer otra salida por donde la materia se escape del Sol. En muchos fenómenos transitorios, bucles inicialmente cerrados pierden estabilidad, ascienden y se dilatan hasta el punto de que sus líneas se abren, expulsando materia hacia el espacio. Recientes observaciones han revelado también la existencia de "balas coronales": nudos de materia densa y fría que se aceleran rápidamente a través de la corona y contribuyen a la salida de materia del Sol.

El campo magnético terrestre protege la Tierra del impacto directo del viento solar. Pero éste distorsiona el campo. En la dirección que mira al Sol, el campo magnético terrestre está comprimido; en la dirección de alejamiento, el campo se estira y forma una larga cola. La región donde el campo magnético terrestre está confinado por el viento solar se llama magnetosfera. Como la propia corona, las magnetosferas de la Tierra y otros planetas constituyen óptimos laboratorios donde estudiar las interacciones entre plasmas y campos magnéticos.

Una magnetosfera es un obstáculo en el camino del viento solar. Fluye éste más deprisa que la velocidad del sonido o de las ondas magnetohidrodinámicas; ninguna perturbación viajará, pues, corriente arriba desde la magnetosfera para ajustar el viento solar al obstáculo. Es la misma situación que cuando un avión vuela a mayor velocidad que el sonido o una lancha se mueve más deprisa que las olas del mar. En todos estos casos, el fluido se ajusta brusca y violentamente al obstáculo, en vez de hacerlo gradualmente, como lo haría si su velocidad fuese inferior a la del sonido o alguna otra clase de onda. El brusco ajuste es una onda de choque, en la que las propiedades del fluido (presión, densidad y velocidad) cambian drásticamente. La onda creada por un avión supersónico es el estampido sónico; la de la lancha es la onda de proa. Un enorme arco de choque está situado en el espacio delante de la magnetosfera terrestre, a unos 60.000 kilómetros de la Tierra. Los vehículos espaciales que crucen el arco de choque pueden recoger valiosa información acerca de la estructura de las ondas de choque en los plasmas.

Si el viento solar fuera constante, la magnetosfera y el arco de choque permanecerían en posiciones fijas determinadas por un equilibrio entre la presión total del viento solar incidente y la presión del campo magnético terrestre. Pero el viento no es más que una prolongación de la corona, fuertemente variable en el espacio y en el tiempo. Al girar el Sol, las irregularidades espaciales de la corona –aquí un agujero coronal, allí un bucle magnético– se traducen en variaciones temporales del viento solar a la distancia de la Tierra. Los sucesos dinámicos del Sol, fulguraciones fotosféricas o transitorios coronales por ejemplo, producen también variaciones en el viento solar.



12. DESARROLLO DE UN BUCLE CORONAL TRANSITORIO, según un modelo matemático ideado por el autor. Las curvas representan posiciones sucesivas de un bucle de materia coronal inmerso en un campo magnético y en proceso de dilatación. La cima del bucle asciende a medida que sus patas se separan, en respuesta quizás a un suceso violento como una fulguración solar o como resultado de movimientos caóticos del gas en la superficie del Sol. El modelo predice que la velocidad de ascenso del bucle es progresivamente mayor que la separación lateral de sus patas. Este resultado indica una situación inestable en la que perturbaciones no muy intensas de la superficie del Sol pueden iniciar rápidas variaciones en la estructura de la corona. El modelo incluye los efectos de la presión gaseosa, el campo magnético y la gravedad. Para hacer el modelo matemáticamente accesible, se supuso, por convención, que la interacción magnética quedaba confinada a delgadas capas coincidentes con las curvas trazadas.

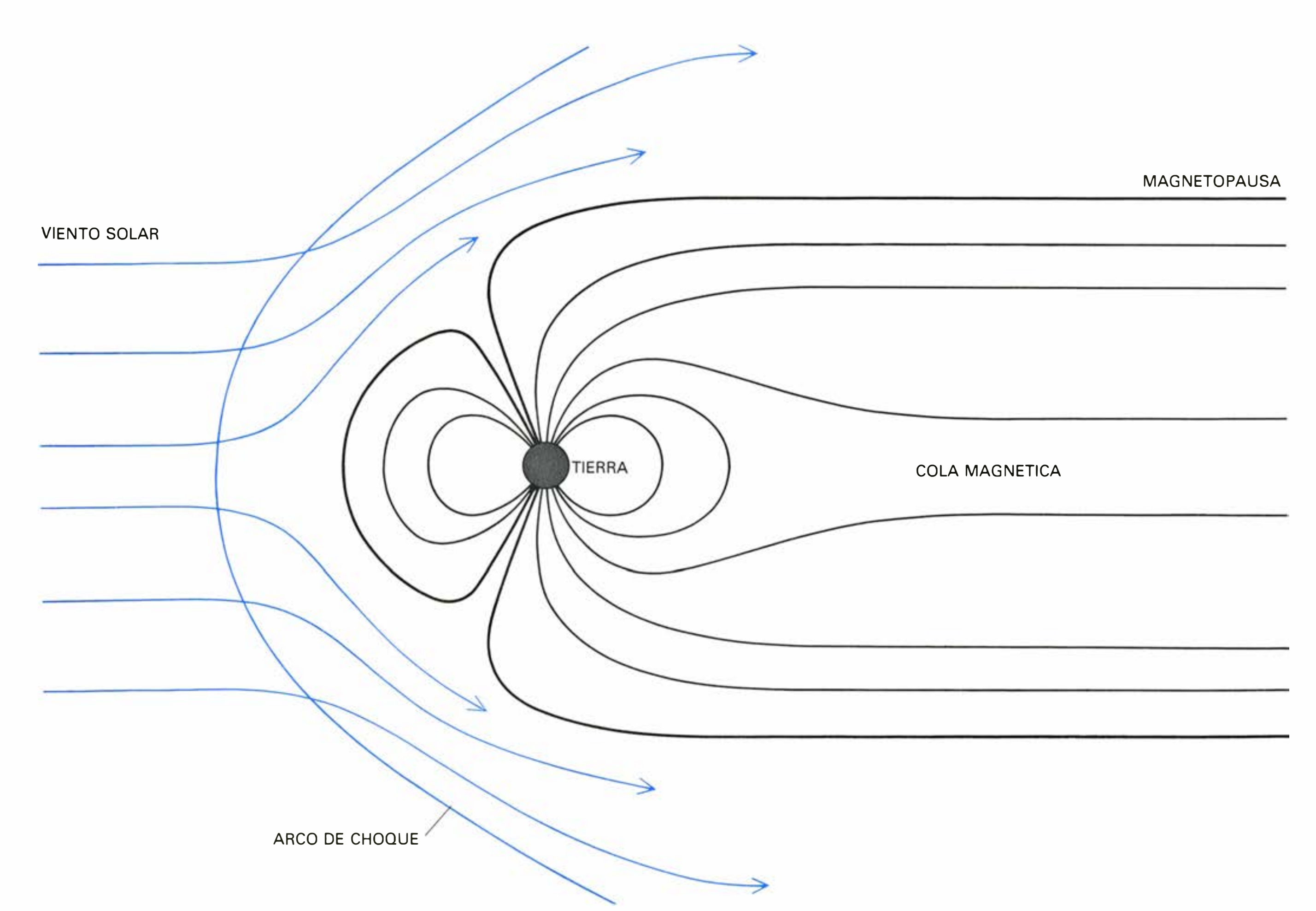
En respuesta a la presión cambiante del viento solar, cambian de posición el arco terrestre de choque y la magnetosfera, comprimiendo el campo magnético terrestre y alterando su intensidad. Las variaciones en el campo magnético inducen corrientes en la ionosfera o capa conductora de la alta atmósfera. Esto produce, a su vez, brus-

cas perturbaciones en la radiocomunicación a gran distancia. Corrientes semejantes pueden inducirse en las líneas telefónicas y de alta tensión, tanto aéreas como subterráneas, y en la Tierra misma. Las perturbaciones intensas del Sol provocan importantes alteraciones en las brújulas, comunicaciones e incluso transmisión de energía eléctrica. En tales períodos, las partículas de alta energía descienden en trayectorias helicoidales a lo largo de las líneas de fuerza magnética a altas latitudes y originan las auroras.

Los chorros de viento solar de alta velocidad emergen de los agujeros coronales. A medida que el viento de alta velocidad empuja al que le antecede, más lento, la zona de transición entre el viento lento y el rápido se comprime y estrecha y la propia transición se torna más brusca. Hacia la época en que el viento solar alcanza la órbita de Júpiter, cinco veces más alejada del Sol que la terrestre, las transiciones se han convertido en ondas de choque tan marcadas que caen cual martillazos sobre el campo magnético de Júpiter. La magnetosfera de este planeta llena un volumen muy grande; si pudiéramos verla, desde la lejanía en que nos hallamos se nos antojaría mayor que el Sol. En respuesta al machaqueo del viento solar, su tamaño puede aumentar tres veces o reducirse a un tercio en cuestión de días. Los vehículos espaciales Pioneer y Voyager, al pasar cerca de Júpiter, cruzaron el arco de choque y la magnetopausa del planeta repetidas veces, lo que indica que las dos estructuras se mueven rápidamente hacia atrás y hacia delante, hinchándose y contrayéndose.

Si la corona aparece todavía en Júpiter, ¿dónde acaba? ¿Cuál es el dominio total de la influencia del Sol? Debe haber un momento en que la presión decreciente del viento solar sea menor que las minúsculas presiones del gas interestelar y del campo magnético de la galaxia. Esta región de transición marca la heliopausa: el final de la extensión física del Sol. Cálculos teóricos realizados diez años atrás sugerían que la heliopausa podría estar entre las órbitas de Júpiter y Saturno. Con vehículos espaciales activos cerca de la órbita de Neptuno, no se ha encontrado signos de la heliopausa. Se cree que el viento solar se prolonga considerablemente más allá de los límites planetarios del sistema solar.

Allende la heliopausa hay un vasto arco de choque causado por el movimiento del sistema solar en bloque a través del medio interestelar. Impulsada por la variabilidad de la corona solar, la heliosfera, región interior de la heliopausa, puede dilatarse y contraerse como un globo. Todo, desde el horno termonuclear del interior solar hasta la heliopausa de su parte exterior –pasando por la zona convectiva, la fotosfera, la cromosfera, la corona y el viento solar– es una componente del Sol. Los seres humanos ocupan una posición privilegiada dentro de este gigantesco objeto celeste, suficientemente lejos del núcleo para vivir cómodamente, pero suficientemente cerca para poder comenzar a comprender la complicada dinámica de una corona dilatada o extensa.



13. LA PRESION DEL VIENTO SOLAR confina el campo magnético terrestre a la región llamada magnetosfera, que se comprime en la región frente al Sol y se alarga formando una cola en la región alejada del Sol. La magnetopausa, o superficie frontera entre el viento solar y la magnetosfera, actúa de barrera e impide que la mayor parte de las partículas del viento lleguen a la Tierra. Sin embargo, cerca de los polos terrestres, las partículas se pueden mover a lo largo de las líneas del campo magnético hasta dentro de la alta atmósfera, dando lugar a las auroras. Delante de la magnetosfera hay un arco de choque que es análogo a la onda de choque que genera el estampido sónico de un avión supersónico. (Esquema realizado por Allen Beechel.)

Computador óptico

El computador que se apoye en el haz de luz podría ejecutar un billón de operaciones por segundo. Se ha construido ya el componente crucial: un análogo óptico del transistor

Eitan Abraham, Colin T. Seaton y S. Desmond Smith

El desarrollo experimentado por el computador digital a lo largo de los últimos 40 años está íntimamente ligado al progreso de la técnica electrónica, hasta el punto de que si preguntamos cómo definir un computador la mayoría de la gente nos contestará que se trata de una máquina de naturaleza electrónica. La verdad es que las operaciones que lleva a cabo son de orden lógico y aritmético; y pueden realizarse de formas muy varias. Desde mediados del decenio de 1970 fue tomando cuerpo la viabilidad de un computador donde las señales se transmitieran a través de haces de radiación láser, en vez de corriente eléctrica. Había un incentivo poderoso para aplicarse a la creación de semejante ingenio óptico: operaría 1000 veces más rápido que un ordenador electrónico.

Los componentes fundamentales de cualquier computador digital son los interruptores capaces de tomar dos estados de transmisión. La velocidad de los cálculos del computador está limitada por el tiempo que requiere un interruptor en cambiar de estado, entre otros factores. En un computador electrónico, la mayoría de sus interruptores son transistores; ahora bien, ni siquiera los transistores más rápidos de cuantos se conocen cambian de estado en un tiempo menor que aproximadamente un nanosegundo, una milmillonésima de segundo. Un aparato óptico análogo al transistor podría pasar de un estado de transmisión al otro en un picosegundo, una billonésima de segundo.

Hemos fabricado una versión experimental de un transistor óptico, accionado por un pequeño cambio de intensidad en el haz incidente de radiación láser. El transistor óptico, que llamaremos transfasor, se funda en una propiedad de ciertos cristales. El incremento en la intensidad de luz provoca un cambio en el índice de refracción del cristal, que mide en qué proporción la luz reduce su velocidad al atravesar un material. Con una combinación adecuada de material cristalino y longitud de onda del láser, podemos aprovechar el cambio de índice de refracción para construir un dispositivo donde la intensidad de la radiación transmitida aumente bruscamente ante una levísima modificación de la intensidad del haz incidente. El modelo experimental alcanza tiempos de conmutación de escasos picosegundos.

El transistor óptico tendría su lugar en la construcción de computadores que procesaran información según el esquema general de las máquinas electrónicas, aunque mucho más deprisa. En un futuro más lejano, las posibilidades de ese transistor transformarían, por sí mismas, la organización del computador. Cada transistor óptico podría ser el lugar de verificación de muchas operaciones de conmutación simultáneas llevadas a cabo a partir de haces láser paralelos; un dispositivo electrónico opera sobre una única señal en un tiempo. Además, los cristales con los que hemos trabajado pueden seleccionar sucesivos niveles de potencia transmitida con sucesivos incrementos en la potencia del haz incidente; por contra, el transistor electrónico del tipo empleado en los computadores tiene sólo dos estados energéticos de salida. La adopción de dispositivos con más de dos estados estables conduciría, seguro, a un nuevo sistema de computador lógico.

El diseño de un computador óptico requeriría múltiples elementos de circuito, amén del transistor óptico. En nuestra labor experimental del laboratorio de la Universidad de Heriot-Watt, pusimos de manifiesto muchos de los elementos necesarios para fabricar un circuito óptico integrado. Habrá que vencer graves dificultades técnicas antes de producir el primer circuito de esos. Lo que no impide que el computador óptico aparezca como un proyecto fascinante de un futuro próximo.

Las tres operaciones básicas de un computador –aritméticas, lógicas y de almacenamiento de información o memoria– se llevan a cabo por dispositivos que tienen dos estados estables. En las operaciones aritméticas, los dos estados representan el 0 y el 1 del sistema binario de numeración. En la evaluación de las proposiciones lógicas, los dos estados representan lo verdadero y lo falso. La memoria del computador almacena los resultados de las operaciones aritméticas y lógicas en un dispositivo que ocupa uno de los dos estados.

Funciones del interruptor

Con el sistema algebraico binario, un computador puede evaluar la verdad de las proposiciones utilizando sólo tres funciones lógicas: la función Y, la función O y la función NO. En la función Y, se considera que una proposición es verdadera si todas sus componentes son verdaderas. En la función O, sólo si alguna de sus componentes es verdadera. En la función NO se invierte el valor de la proposición. Las operaciones lógicas más complicadas se pueden construir a partir de estas tres funciones básicas; así se llevan a cabo operaciones aritméticas como la suma.

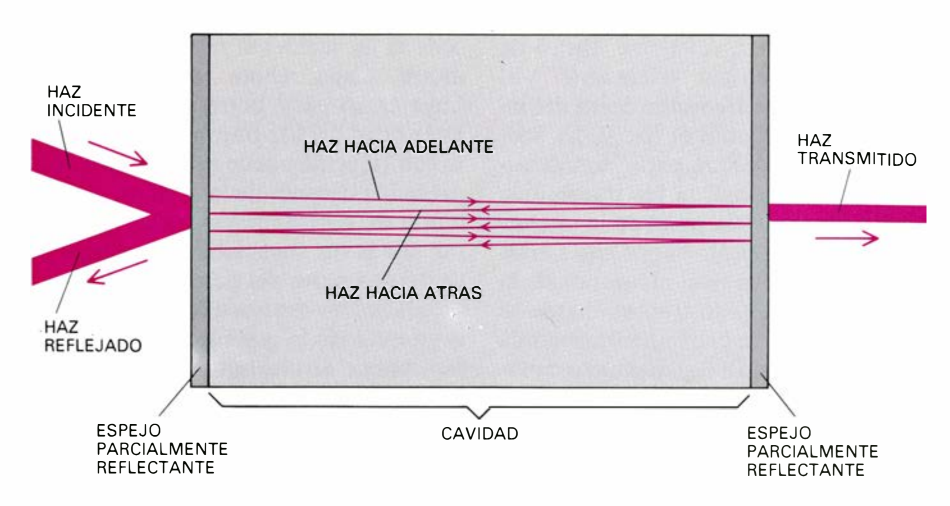
El ordenador requiere, pues, un dispositivo que represente los valores 0 y 1, o verdadero y falso, de forma física y pueda ensamblarse en dispositivos a gran escala que realicen las tres funciones lógicas. Obviamente, el dispositivo debe tener dos salidas, fácilmente diferenciables, una para cada estado. Si el interruptor trabaja a gran velocidad, el tiempo necesario para conmutar entre los dos estados debe ser pequeño. Otras consideraciones sugieren que el interruptor más apreciado sea el más pequeño, el de más fácil fabricación y el que requiere menor potencia.

El primer conmutador electrónico práctico fue el tubo de vacío. Macizo y

lento, su funcionamiento generaba gran cantidad de calor, lo que limitaba su vida operativa. Resultaba así que los computadores basados en el tubo de vacío eran enormes y su capacidad operativa, pequeña. Las máquinas más potentes requerían la sustitución incesante de los tubos defectuosos.

La invención del transistor en 1947 produjo un interruptor más pequeño, rápido y eficaz. En su forma más simple, el transistor consta de tres placas paralelas de materiales semiconductores. (Por semiconductor se entiende un material cuya conductividad eléctrica es intermedia entre la de un conductor y un aislante; sus propiedades eléctricas se alteran con facilidad.) Las capas externas reciben los nombres de emisor y colector; de base, la intermedia. Las propiedades eléctricas de las capas se escogen de suerte tal que cualquier cambio levísimo en la corriente que va desde la base hasta el colector ocasione una modificación mucho mayor en la corriente que va del emisor al colector.

Para representar una gran corriente desde el colector se emplea el 1; para simbolizar otra pequeña, el 0. Por combinación de los transistores y otros componentes se consiguen estructuras que ejecuten las funciones Y, O y NO; estos circuitos se denominan puertas lógicas. La puerta Y producirá una corriente intensa si en todas sus entradas hay una corriente elevada. La puerta "O" produce una corriente intensa, si en alguna de sus entradas hay una corriente elevada. La puerta NO produce una corriente elevada si en la entrada hay una corriente baja, y una corriente baja si en la entrada hay una corriente alta.



1. INTERFEROMETRO DE FABRY-PEROT, esquematizado. Constituye el fundamento del interruptor óptico construido por los autores. El dispositivo consta de dos espejos planos colocados en paralelo, con un material transparente dentro del espacio entre ambos. Este espacio se llama cavidad. Cada espejo refleja parte de la luz que incide sobre su superficie, y transmite el resto. Cuando un haz incide sobre el espejo frontal, parte de la luz se refleja y parte pasa al interior de la cavidad. El haz transmitido hacia adelante viaja a través de la cavidad y llega al espejo trasero, donde parte de la luz abandona el dispositivo y parte se refleja formando un haz hacia atrás. El haz continúa rebotando hasta que la luz escapa de la cavidad. Modificando el tamaño de la cavidad, el material de su interior o la longitud de onda de la luz cambia la transmisión del interferómetro. En un interruptor óptico, el material ubicado dentro de la cavidad posee propiedades que permiten que la intensidad del haz incidente controle la transmisión.

Limitaciones del transistor

En razón de su tamaño, material, diseño y potencia, las puertas lógicas proyectadas con semiconductores necesitarán de 1 a 1000 nanosegundos para operar. Desde un supuesto ideal, las puertas más veloces podrían realizar mil millones de operaciones por segundo. En la práctica, la celeridad de proceso es bastante menor. La física nos dice que el actual, cifrado en torno al nanosegundo, se halla muy cerca del tiempo de conmutación mínimo al alcance de un dispositivo semiconductor de conmutación.

Para que un transistor conmute, debe atravesar su base una corriente de electrones. La velocidad de desplazamiento de los electrones por un semiconductor tiene unos límites perfectamente acotados. Estrechando la anchura de la base, se abrevia el tiempo de conmutación; ahora bien, la base no puede estrecharse sin tasa. Y quizá hemos alcanzado ya esos límites. Se podrán lograr, pues, reducciones sustanciales en el tiempo de selección, pero no a través de una modificación o mejora del diseño del transistor, sino en virtud de la adopción de una nueva técnica de conmutación.

Un camino: para transportar la información a través del circuito, sustituir la corriente electrónica por una señal de otro tipo. La máxima velocidad que una señal transportadora de información puede alcanzar es la de la luz; en la luz (o una radiación electromagnética de otra parte del espectro), tenemos, pues, un candidato idóneo para acarrear información. Siguiendo esta línea de razonamiento, hemos utilizado la radiación coherente de un láser para manipular dispositivos conmutadores, dotados de muchas de las capacidades del transistor. Los dispositivos conmutan entre dos estados de salida diferenciados, a pleno rendimiento; sirven de elementos de memoria y, por una adecuada selección de materiales y haces láser, pueden obtenerse puertas lógicas, rápidas y simples.

El punto de arranque del transistor óptico es una pieza de un aparato óptico, ingenioso y de amplio uso que se conoce por interferómetro de Fabry-Perot, inventado por los físicos franceses Charles Fabry y Alfred Perot en 1896. Fabry y Perot lo emplearon para medir la longitud de onda de varios colores del espectro de la luz; posteriormente se ha aplicado a otros muchos usos. El dispositivo de Fabry-Perot, en su forma más simple, consta de dos espejos planos colocados paralelamente y separados por un espacio, que recibe el nombre de cavidad, donde puede insertarse un material transparente a la longitud de onda requerida.

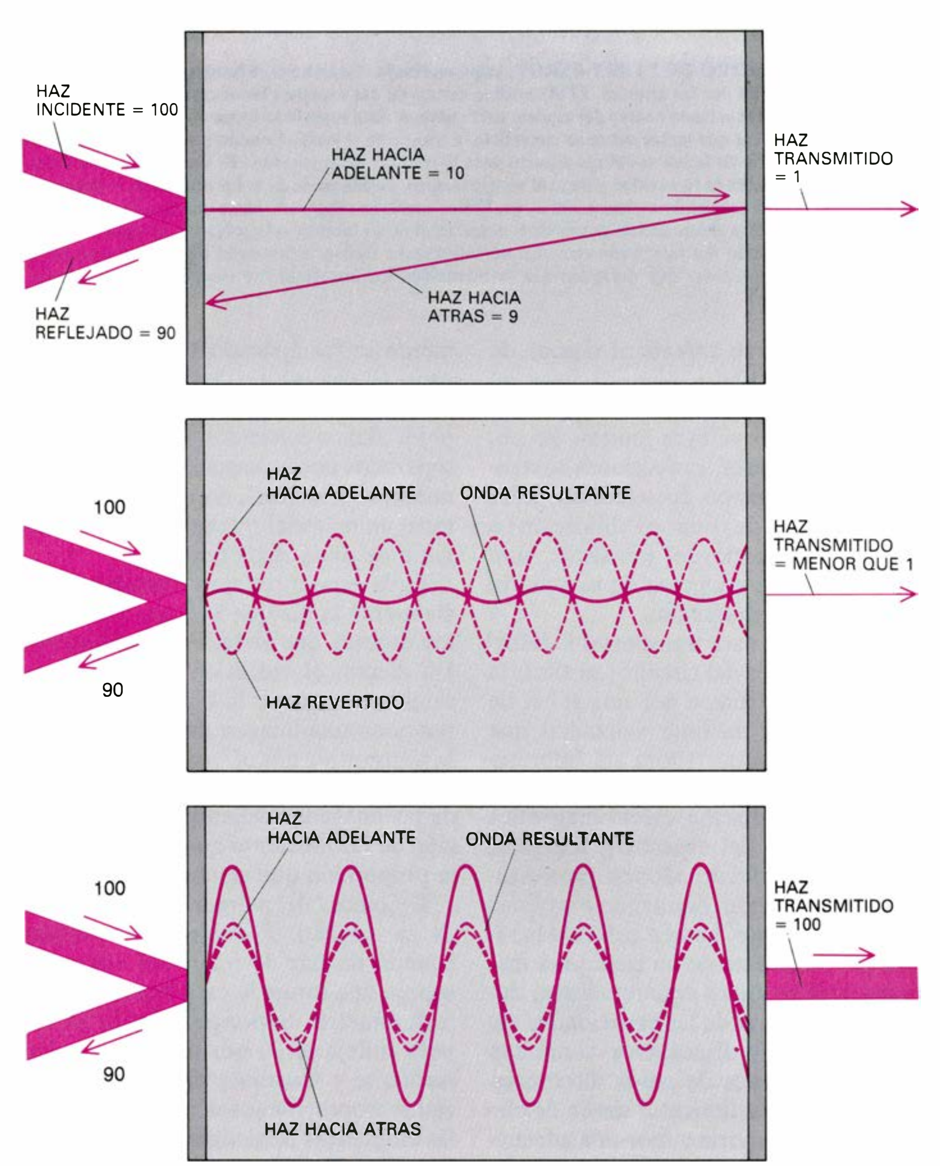
Cada espejo refleja y transmite parcialmente la luz que le llega. Abundan los espejos que reflejan parcialmente. En efecto, el vidrio de un escaparate semeja un espejo: la luz reflejada proporciona una imagen del observador y la transmitida nos ofrece una imagen de los objetos que protege. La intensidad de las imágenes depende de la proporción de luz incidente que se refleja y de la proporción que se transmite.

Dejemos, de momento, el material de la cavidad. Pensemos qué sucede cuando un haz de luz choca contra el espejo que forma la cara frontal del interferómetro. Supongamos que el espejo refleja el 90 por ciento de la luz incidente y transmite el 10 por ciento (tales proporciones son muy próximas a las empleadas para algunos de nuestros trabajos). De este modo, cuando el haz de luz choca con el espejo frontal el 90 por ciento de la misma se refleja y el 10 por ciento se introduce en el interferómetro.

La luz, que ahora tiene la décima parte de su intensidad original, constituye el haz que se propaga "hacia adelante" y pasa al espejo trasero. Las propiedades del espejo trasero son iguales que las del frontal. Y así, en el espejo trasero el 90 por ciento del haz hacia

adelante vuelve a reflejarse dentro de la cavidad, como haz "hacia atrás" y el 10 por ciento se transmite fuera del interferómetro. Como el haz hacia adelante tiene la décima parte de intensidad que el original, la luz transmitida tendrá la centésima parte de la intensidad del haz original. El haz hacia atrás retorna, por supuesto, al interior de la superficie del espejo frontal, donde se divide otra vez en una parte transmitida y otra reflejada. El haz atenuado continúa rebotando dentro de la cavidad, debilitándose cada vez más, hasta que toda la luz incidente sale del interferómetro. Cada rebote sucesivo contribuye en su parte correspondiente a la intensidad del haz transmitido; pero esta contribución puede considerarse despreciable después de la primera vez. A modo de aproximación, cabe considerar que la luz transmitida constituye la centésima parte del haz incidente.

Si las cifras expresadas más arriba recogiesen todo lo que sucede en el interferómetro, el dispositivo carecería de interés a la hora de estudiar un interruptor para un computador óptico. Podríamos aumentar la intensidad del haz transmitido con sólo modificar las propiedades de los espejos. Pero se impone una matización adicional. Los haces hacia delante y hacia atrás no pueden separarse en la realidad con la nitidez con que lo hemos hecho en nuestra sumaria descripción. En un interferómetro, la interacción de los haces condiciona la intensidad de la luz en la cavidad, que, a su vez, determina la transmisión.



2. INTERFERENCIA ENTRE LOS HACES que se propagan hacia adelante y hacia atrás en la cavidad Fabry-Perot. Modifica sensiblemente la intensidad de la luz en su interior y, en consecuencia, la transmisión del dispositivo. Los espejos frontal y trasero se eligen de suerte que reflejen el 90 por ciento de la luz que reciben y transmitan el 10 por ciento. Sin tener en cuenta la interferencia, la situación es la siguiente (*arriba*). El 10 por ciento del haz incidente pasa dentro de la cavidad. Cuando el haz atenuado alcanza el espejo trasero se transmite un haz con un 1 por ciento de intensidad; el haz que se propague hacia atrás será igual al 9 por ciento del original. Además los haces hacia adelante y hacia atrás interactúan como ondas dando lugar a otra onda que, en cada punto, es igual a la suma de las componentes. Si la cresta de un haz se alinea con el valle del otro, lo que se conoce por interferencia destructiva, se forma una onda pequeña (*centro*). La intensidad de la luz es insignificante y la transmisión puede ser inferior al 1 por ciento del haz incidente. Si las dos ondas se alinean en interferencia constructiva, resulta una onda grande (*abajo*). Por haber muchos haces propagándose hacia adelante y hacia atrás, la intensidad en la cavidad puede multiplicar por 10 la del haz incidente. Así, cuando la interferencia es plenamente constructiva, el haz transmitido puede llegar a tener igual intensidad que la ostentada por el haz incidente original.

Interferencia en la cavidad

La luz disfruta de algunas propiedades de las ondas y de otras que caracterizan a las partículas. Según los casos, convendrá una u otra descripción. En este contexto, consideramos más adecuada la descripción ondulatoria. Los haces hacia delante y hacia atrás, junto a los sucesivos rebotes cada vez más débiles, pueden considerarse cual ondas en propagación a través de la cavidad. El haz incidente penetra en el interferómetro perpendicularmente a la cara frontal: de esta forma, todos los haces viajan juntos, siguiendo la misma trayectoria.

Sin olvidar que hay muchos haces reflejados dentro de la cavidad, vamos a centrarnos en la interacción de sólo dos de ellos: el que se propaga hacia adelante y el haz que se propaga hacia atrás. El resultado de esta interacción depende de la fase de las ondas, es decir, de la posición relativa de las crestas y las depresiones. Las crestas se alinean con las crestas y las depresiones con las depresiones en una relación conocida como interferencia constructiva. Alternativamente, las crestas de una onda pueden alinearse con los valles de la otra, en una interferencia destructiva. Caben fases intermedias entre las indicadas.

Por interferencia de dos ondas solapadas se forma una tercera onda. La amplitud de la onda resultante en cada punto de la cavidad será la suma de las amplitudes de las ondas componentes en ese punto. Si la interferencia es constructiva, la amplitud será grande; y pequeña, si destructiva. El efecto de la interferencia en la transmisión del interferómetro de Fabry-Perot no es, ni de lejos, despreciable. En plena interferencia destructiva, la intensidad de la luz en la cavidad es casi cero y la transmisión irrelevante. En plena interferencia constructiva, los múltiples haces hacia adelante y hacia atrás se refuerzan entre sí, de tal modo que la intensidad de la luz en la cavidad multiplica por diez la intensidad del haz incidente

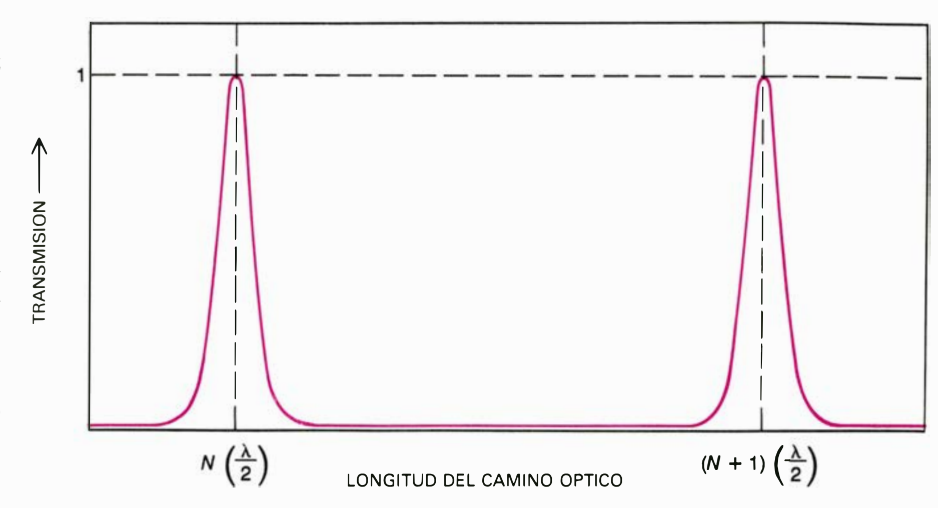
original. Además, como el espejo trasero transmite el 10 por ciento de la luz que recibe, en plena interferencia constructiva el interferómetro transmite un haz igual al incidente.

Efecto del índice de refracción

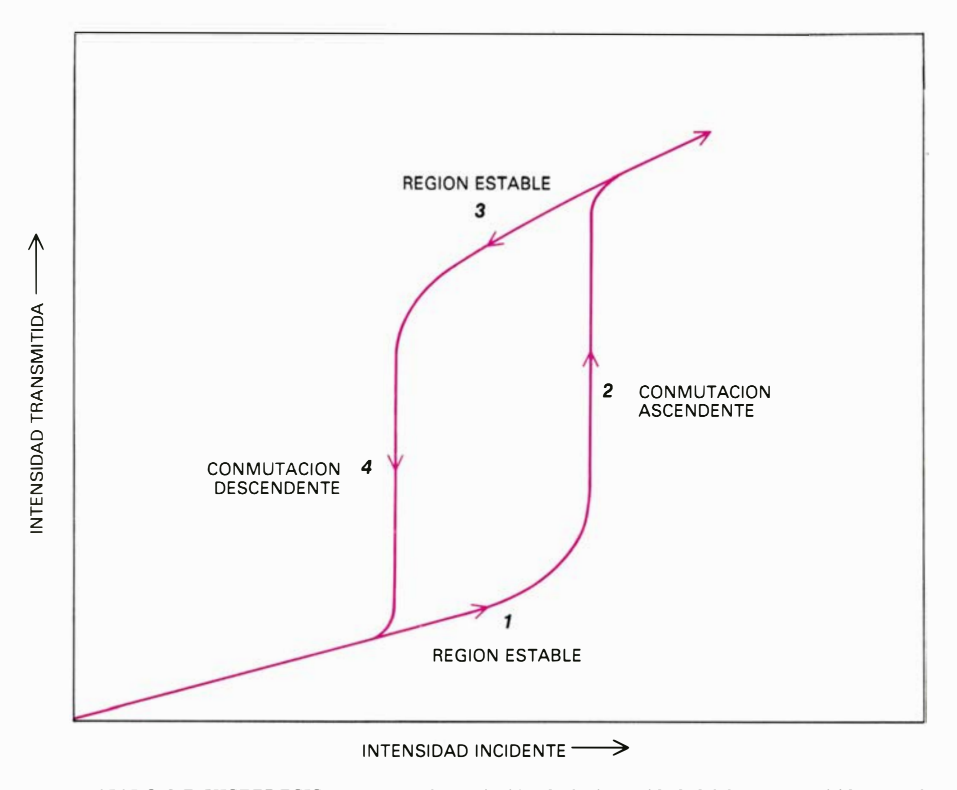
Para ajustar la fase de los haces hacia adelante y hacia atrás, y por tanto la transmisión del interferómetro, se modifica la longitud de la cavidad o la longitud de onda de la luz incidente. Pero no le sirve de mucho al transistor óptico ningún procedimiento de ésos. Antes bien, lo que sí le importa es que la relación de fases puede modificarse también alterando las propiedades ópticas del material en la cavidad, cosa que hasta este momento no hemos tenido en cuenta.

La luz alcanza su máxima velocidad de propagación en el vacío. En otro medio transparente, distinto del vacío, la celeridad de la luz remite en una cuantía que es función del medio. Este freno de la onda luminosa produce el fenómeno de la refracción; el índice de refracción de un material expresa el cociente entre la velocidad de la luz en el vacío y la velocidad de la luz en el material en cuestión. Por ejemplo: si el índice de refracción de cierto tipo de vidrio es de 1,5, esto significa que mientras la luz viaja a 300.000 kilómetros por segundo en el vacío, en el vidrio lo hará a 200.000 kilómetros por segundo. La ralentización de la luz va acompañada de una reducción de la longitud de onda, pero no de un cambio en la frecuencia. Se puede comprobar que incluso sin alterar la longitud de la cavidad ni la longitud de onda de la luz incidente, un cambio en el índice de refracción del material modificará las relaciones de fase de los haces en ella debido a una variación de la longitud de onda. Seleccionando un material con el índice de refracción apropiado, podemos provocar que los haces hacia adelante y hacia atrás interfieran entre sí constructiva o destructivamente.

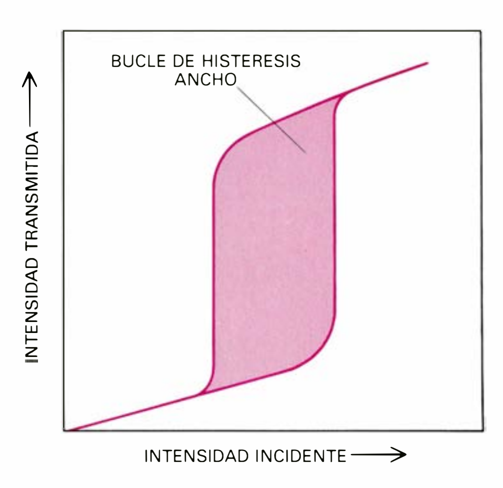
Un parámetro importante en el análisis de un sistema óptico es la denominada longitud del camino óptico: la longitud óptica. En el medio interior de la cavidad viene dada por la longitud física multiplicada por su índice de refracción. Se demuestra que se tiene interferencia totalmente constructiva cuando la longitud óptica equivale a la distancia correspondiente a un número entero de semilongitudes de onda de la luz incidente. (Se llama semilongitud de onda a la distancia que media entre una cresta y un valle.) De esta forma ten-

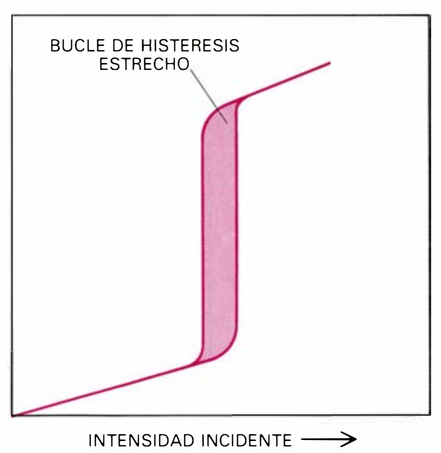


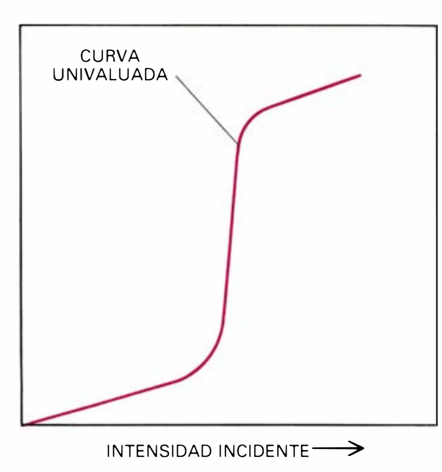
3. FUNCION DE AIRY, para expresar la relación entre la transmisión del interferómetro y la longitud del camino óptico del material de la cavidad. La transmisión es el cociente entre la intensidad del haz incidente y la del transmitido; su máximo es 1, que lo alcanza cuando la interferencia dentro de la cavidad es enteramente constructiva. Se llama longitud del camino óptico a la longitud física del material de la cavidad multiplicada por su índice de refracción (el índice de refracción mide el grado de deceleración de la luz al atravesar una sustancia transparente). La interferencia es plenamente constructiva cuando la longitud del camino óptico vale un número entero de semilongitudes de onda (λ) de la luz incidente.



4. EL CICLO DE HISTERESIS representa la variación de la intensidad del haz transmitido por el interferómetro en función de la intensidad del haz incidente. En la cavidad hay un material con un índice de refracción no lineal: varía según la intensidad de la luz en el material. Los cambios de intensidad en el haz incidente modifican el índice de refracción y la longitud óptica y, por ende, la transmisión. Si la intensidad del haz incidente aumenta despacio, la intensidad transmitida en un comienzo ofrece un aumento gradual (*1*). El cambio del índice de refracción va acercando poco a poco la longitud del camino óptico al del valor correspondiente a la interferencia constructiva, con lo cual aumenta la intensidad de la luz en la cavidad; cualquier cambio en la intensidad alterará el índice de refracción. En un punto determinado, esta realimentación positiva se hace muy fuerte. Alcanzado un umbral, la transmisión aumenta rápidamente (*2*). Si entonces reducimos un poco la intensidad incidente, la transmisión no cae al valor original porque todavía se conserva bastante luz en la cavidad para mantener el índice de refracción y la longitud del camino óptico próximos a los valores correspondientes a la transmisión máxima; la transmisión desciende gradualmente (*3*). Existe un segundo umbral, a partir del cual una reducción de intensidad del haz incidente modifica el índice de refracción lo suficiente para producir una drástica pérdida de intensidad en la cavidad; la transmisión cae entonces rápidamente hasta el bajo nivel inicial (*4*). De los dispositivos ópticos que presentan ciclos de histéresis se dice que son ópticamente biestables. Las regiones *1* y *3* de la curva son zonas estables, donde la transmisión varía muy poco con los cambios de intensidad.







5. MODIFICACION DE UN BUCLE DE HISTERESIS en un dispositivo óptico biestable mediante la variación de la longitud del interferómetro, la longitud de onda del rayo incidente o el material dentro de la cavidad. En el ciclo de histéresis fundamental, las regiones estables son anchas y el lazo muy abierto (*izquierda*). Un nivel de intensidad incidente puede corresponder a uno o dos niveles de intensidad transmitida. Regulando el interferómetro se pueden reducir las regiones estables y estrechar el lazo (*centro*). El límite del proceso es una curva univaluada, en la que cada nivel de intensidad incidente corresponde a un solo nivel de intensidad transmitida (*derecha*). Esta curva es similar a la característica de transferencia de un transistor electrónico; en el transistor óptico se utiliza un interferómetro que presenta la referida curva de transmisión. Del ciclo de histéresis y sus principios se ocupa la figura 4.

dremos muchas longitudes ópticas posibles que den lugar a la interferencia totalmente constructiva. Cuando la longitud óptica queda a la mitad del espacio entre las distancias correspondientes a múltiplos enteros de la semilongitud de onda, el resultado será una interferencia totalmente destructiva.

La intensidad de la luz transmitida por el interferómetro varía en función de las relaciones de fase de los haces; por tanto, la intensidad transmitida varía con la longitud óptica. La función que describe la relación entre la longitud óptica y la transmisión del interferómetro (definida como el cociente de la intensidad del haz transmitido respecto de la del incidente) se conoce como función de Airy; reviste especial interés en la construcción de un transistor óptico. En los dispositivos Fabry-Perot que hemos fabricado, la función tiene un máximo destacado para cada punto donde la longitud óptica corresponde a un número entero de semilongitudes de onda; entre máximos, el valor de la función es pequeño y sólo cambia de un modo gradual. De esta forma, a medida que la longitud óptica se aproxima al valor correspondiente a la transmisión máxima, la transmisión crece poco a poco, hasta alcanzar un punto umbral, a partir de cuyo momento aumenta con rapidez.

El efecto ejercido por el índice de refracción en la longitud óptica implica que la transmisión de una cavidad puede situarse en un nivel prefijado; basta, para ello, con inserir dentro de la misma un material dotado del índice de refracción apropiado. Cambiando un material por otro, se modificará también la transmisión. Resulta evidente que un procedimiento de conmutación de este estilo resultaría muy poco útil para un computador. A pesar de ello, en el siglo XIX, se hubiera considerado la única forma de modificar el índice de refracción. James Clerk Maxwell, en su trabajo fundamental sobre la radiación electromagnética, postuló que la refracción y otras interacciones de la luz con la materia eran independientes de la intensidad de la luz. Si ello fuese siempre cierto, el índice de refracción se mantendría constante y el rendimiento del interferómetro sería directamente proporcional a la intensidad del rayo incidente. Bajo determinadas condiciones, muchas sustancias valdrían para este fin; serían "lineales" con respecto al índice de refracción.

Refracción no lineal

Cuando el desarrollo del láser puso en nuestras manos una fuente más potente de radiación coherente (esto es, una radiación en la que todas las crestas y valles están alineados), quedó claro que no todos los materiales eran lineales con respecto al índice de refracción. Al concentrar la radiación láser sobre determinadas sustancias, la refracción varía con la intensidad del haz. Estos materiales son no lineales con respecto al índice de refracción. Con una sustancia no lineal en la cavidad Fabry-Perot, modificaremos el índice de refracción sin más que cambiar la intensidad del haz incidente.

Este efecto, muy sutil, merece nuestra atención. Imaginemos una cavidad de longitud fija en cuyo interior haya un material no lineal; dirijamos un haz láser, de intensidad ajustable, hacia el interior. La intensidad inicial del haz provoca que el índice de refracción produzca una longitud óptica en la región de la curva de Airy entre máximos; la transmisión es baja. Si potenciamos paulatinamente la intensidad del haz, el índice de refracción y la longitud óptica cambiarán poco a poco; la transmisión crecerá moderadamente. Llegará un momento en que el índice de refracción y la intensidad de la luz de la cavidad produzcan un bucle de realimentación positiva, esto es, se reforzarán mutuamente. El cambio del índice de refracción trae consigo que los haces se acerquen más a la interferencia constructiva, con lo que aumenta la intensidad en la cavidad; la intensidad aumentada modifica a su vez el índice de refracción. En un cierto nivel crítico, el efecto de realimentación regenerativo llega a ser muy fuerte. Se alcanza la región máxima de la curva de Airy y la transmisión, en efecto, crece aceleradamente.

En el punto máximo de la función de Airy, el cociente de la intensidad transmitida respecto a la incidente es 1. Cuando la intensidad del haz láser baja un poco del valor máximo, la transmisión de la cavidad no vuelve a caer a su bajo valor inicial. La razón estriba en que la intensidad de la luz en la cavidad es suficiente para mantener el índice de refracción y la longitud óptica próximos a los valores que producen la máxima transmisión. A medida que remite apresuradamente la intensidad del haz incidente, la transmisión decrece sólo de forma gradual. Llegados a cierto valor, el índice de refracción y la intensidad en la cavidad se atenúan mutuamente. Una ligera relajación de la intensidad incidente modifica el índice de refracción lo suficiente para reducir sustancialmente la intensidad en la ca-

vidad. Se produce así una caída brusca de la transmisión.

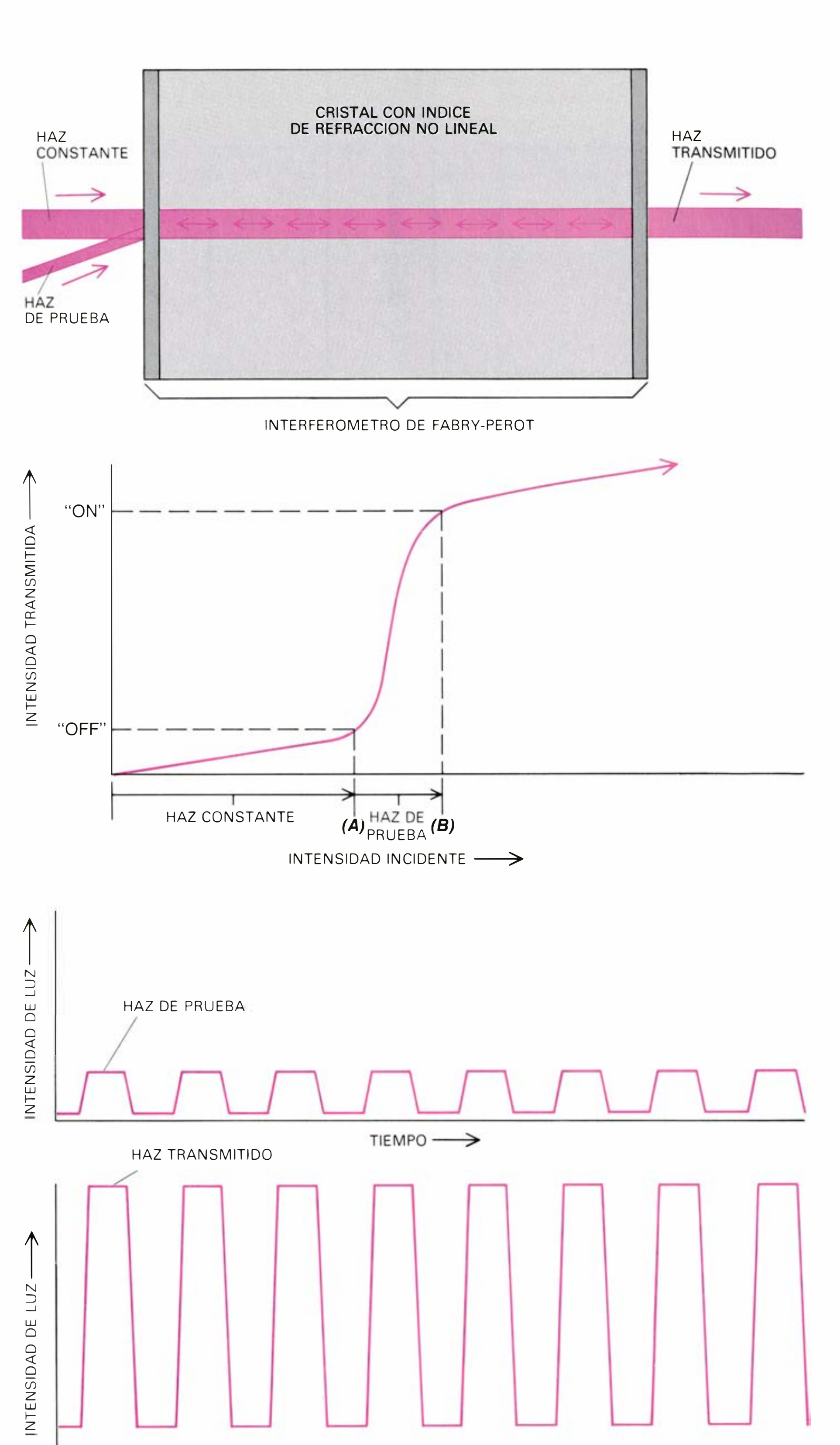
Si se dibujase la intensidad de la radiación transmitida en función de la intensidad del haz incidente, la curva resultante sería un lazo con cuatro secciones [*véase la figura 4*]. La curva arranca con una intensidad incidente baja, y asciende suavemente; hasta un fuerte incremento en la intensidad incidente produce sólo un aumento leve de la transmisión. En el umbral, la curva se transforma bruscamente: un ligero aumento en la intensidad provoca un brusco incremento en la transmisión. El retorno al punto de partida sigue un camino diferente. En un comienzo, la curva se inclina hacia abajo con suavidad; aunque remita la intensidad, la transmisión se mantiene todavía elevada. Pero al llegar a un segundo punto umbral, la curva cae en picado: una ligera pérdida de intensidad atenúa grandemente la transmisión.

El bucle formado constituye lo que se denomina un ciclo de histéresis. De un dispositivo óptico que presente este ciclo se dice que es ópticamente biestable, porque dispone de dos zonas estables, donde la intensidad transmitida varía poco con la intensidad incidente. Habrá que tener en cuenta que, en un dispositivo biestable, un valor particular de la intensidad incidente, podrá asociarse con dos niveles de intensidad transmitida; la transmisión real depende de las condiciones previas de la cavidad.

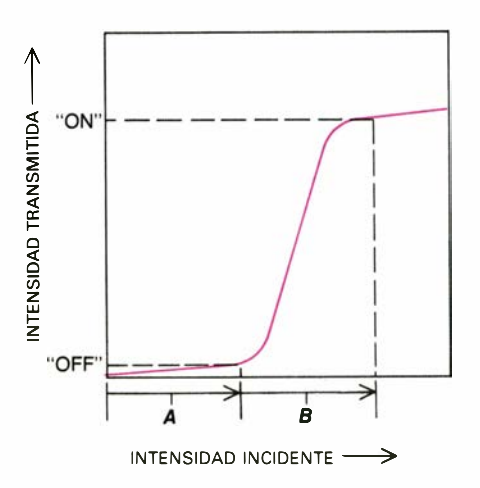
Dispositivos ópticos biestables

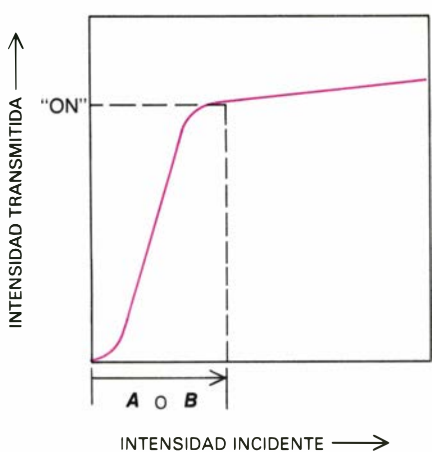
La existencia de la biestabilidad óptica fue predicha por Abraham Szöke y sus colaboradores del Instituto de Tecnología de Massachusetts en 1969. Los primeros en observarla fueron Hyatt M. Gibbs, Samuel L. McCall y Thirumalai N. C. Venkatesan, de los laboratorios Bell, en 1976. Desde esa fecha, los dispositivos ópticos biestables han sido objeto de trabajo en muchos centros.

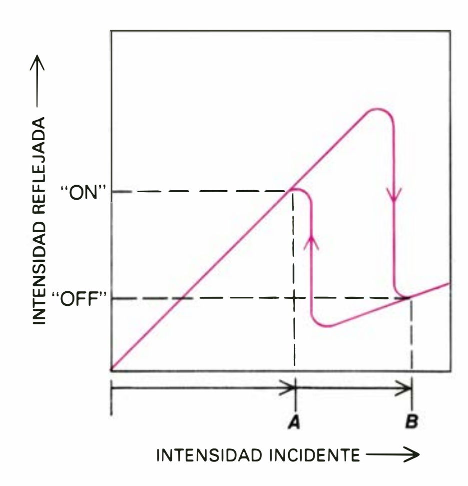
Al preparar un dispositivo óptico biestable, importa elegir un material que posea una sustancial no linealidad en su índice de refracción. En nuestro ensayo utilizamos el semiconductor compuesto antimoniuro de indio, que presenta una acusada respuesta no lineal. Como la mayoría de los semiconductores, el antimoniuro de indio es opaco a la radiación en la parte visible del espectro, pero es transparente a la radiación infrarroja de algunas longitudes de onda. Nuestra fuente de radiación es un láser de monóxido de carbo-



6. TRANSISTOR OPTICO ideado por los autores. Se funda en dos haces láser ajustados con precisión y enfocados hacia la cara frontal de un interferómetro Fabry-Perot, que contiene una sustancia con un índice de refracción no lineal (*arriba*). El haz constante es potente e invariable. El haz de prueba es débil y puede ser modulado. La intensidad del haz constante (*A*) se elige de manera que proporcione una intensidad transmitida baja que se encuentre por debajo de la zona abrupta de la curva de transmisión univaluada (*centro*). A la hora de elegir la intensidad del haz de prueba (*B*) se mira que, cuando se sume a la intensidad del haz constante, la transmisión se eleve por encima de la zona abrupta de la curva. La baja transmisión o estado desconectado ("off ") puede representar el valor 0 en la lógica binaria del computador; la alta transmisión o estado conectado ("on"), el valor 1. En virtud de la pendiente abrupta de la curva de transmisión, basta un ligero cambio en el haz de prueba para ocasionar grandes alteraciones en la intensidad transmitida (*abajo*). En este aspecto, los transistores ópticos y electrónicos son análogos.







7. PUERTAS LOGICAS OPTICAS construidas a partir de transistores ópticos. Las puertas lógicas son dispositivos utilizados en un computador digital para realizar operaciones aritméticas y evaluar la verdad de las proposiciones. Se requieren tres tipos de puertas. Aquí indicamos los principios en razón de los cuales se han realizado las versiones ópticas. La puerta "Y" alcanza una elevada energía de salida en el caso exclusivo en que todas sus entradas están activas (*izquierda*). Si los dos haces láser incidentes se eligen de forma que cada uno alcance una intensidad transmitida justo por debajo de la zona abrupta de la curva de transferencia, tendremos una puerta "Y": ambos haces incidentes deben estar presentes para conseguir una elevada transmisión. La puerta "O" alcanza una elevada energía de salida si alguna de sus entradas es elevada (*ilustración central*). Si los haces incidentes están ajustados de manera que cada uno alcance una transmisión por encima de la zona abrupta, el resultado es una puerta "O": cada haz incidente está capacitado para rendir una elevada transmisión. La puerta "NO" debe ofrecer una elevada energía de salida cuando la de entrada es baja, y viceversa. La puerta óptica "NO" hace uso de la luz reflejada desde el interferómetro (*esquema de la derecha*). El haz reflejado se asocia al bucle de histéresis. Por razón de poseer la misma forma del bucle, una energía de entrada elevada (*A*) proporciona una baja de salida, mientras que una baja energía de entrada (*B*) rendirá una salida elevada.

no cuya longitud de onda puede ajustarse dentro de una estrecha zona del infrarrojo.

Por interferómetro usamos un cristal pequeño y rectangular de antimoniuro de indio. Los bordes del cristal suelen medir escasos milímetros de largo; preparamos también dispositivos agrupando muchos cristalitos. En algunos casos, las caras frontal y trasera del cristal se revestían con una sustancia reflectora, que hacía las veces de espejo; este recubrimiento no es imprescindible, ya que la cara pulimentada del mismo cristal sirve de espejo por sí misma. Un dispositivo situado detrás del espejo trasero mide la intensidad del haz transmitido.

A escala atómica, hay muchos mecanismos que modifican el índice de refracción de un semiconductor. En el antimoniuro de indio el mecanismo es complejo y aún requiere ulteriores investigaciones. Se conocen ya algunos aspectos del proceso y lo que se expone a continuación es un resumen simplificado de lo que se sabe con certeza.

La interacción entre la radiación electromagnética y un material sólido es mediatizada casi enteramente por los electrones en el sólido; en las longitudes de onda de la luz visible o de la radiación infrarroja, son los electrones externos de cada átomo, principalmente, los que participan. En un buen conductor eléctrico, así un metal, los electrones externos gozan de libertad para moverse por el material; en un buen aislante, cada electrón está ligado estrechamente a cada átomo en particular y, para liberarlo, se ha de invertir un elevado coste energético. La situación en un semiconductor es intermedia entre estos extremos; en efecto, cada electrón externo está ligado a un átomo, pero sólo se necesita una pequeña cantidad de energía para que el electrón quede libre.

Estructura de bandas de un semiconductor

Estas diferencias en la configuración electrónica, que determinan las propiedades ópticas y eléctricas del material, se entenderán mejor si nos fijamos en los niveles energéticos asequibles a cada electrón. En un átomo aislado, los niveles están nítidamente definidos; no así en un sólido, donde los niveles asociados resultantes de la suma de muchos átomos forman, como quien dice, amplias bandas de energía posibles, separadas por intervalos prohibidos, que representan energías que ningún electrón puede alcanzar. Si una banda está ocupada del todo, significa que cada nivel de la misma está lleno y cada electrón en la banda está confinado a la vecindad de un átomo. Esto se explica porque, para poner un electrón en movimiento, es necesaria una pequeña cantidad de energía extra y, en este caso, no hay ningún nivel de la banda al cual pueda llegar un electrón que adquiera dicha energía extra. En el caso de una banda parcialmente ocupada, un electrón que adquiera dicha energía adicional se moverá hasta un nivel ligeramente superior a la banda, alcanzando así su libertad.

Los metales tienen una banda parcialmente ocupada. En un aislante, la banda más alta ocupada está completamente llena (la banda de valencia) y queda separada por un amplio intervalo prohibido de la siguiente banda permitida (la banda de conducción), que está vacía. Un electrón podría ponerse en movimiento en un aislante en el caso exclusivo de que absorbiera bastante energía para cruzar el intervalo prohibido, lo cual es muy poco probable. Un semiconductor tiene también completa la banda de valencia y desocupada la banda de conducción, pero el intervalo prohibido entre ambas bandas es más pequeño.

Son diversos los caminos por los que un electrón, en un semiconductor, puede ganar la pequeña cantidad de energía necesaria para ir desde uno de los niveles más altos de la banda de valencia hasta uno de los niveles más bajos de la banda de conducción. Cuando un electrón es impulsado a la banda de conducción deja tras de sí un "hueco" en la banda de valencia; éste actúa en algunos aspectos como una entidad con carga positiva igual en magnitud a la carga negativa del electrón. Cuando el electrón anterior retorna a la banda de valencia se sitúa en un hueco, que de esta forma desaparece; dicho proceso recibe el nombre de recombinación.

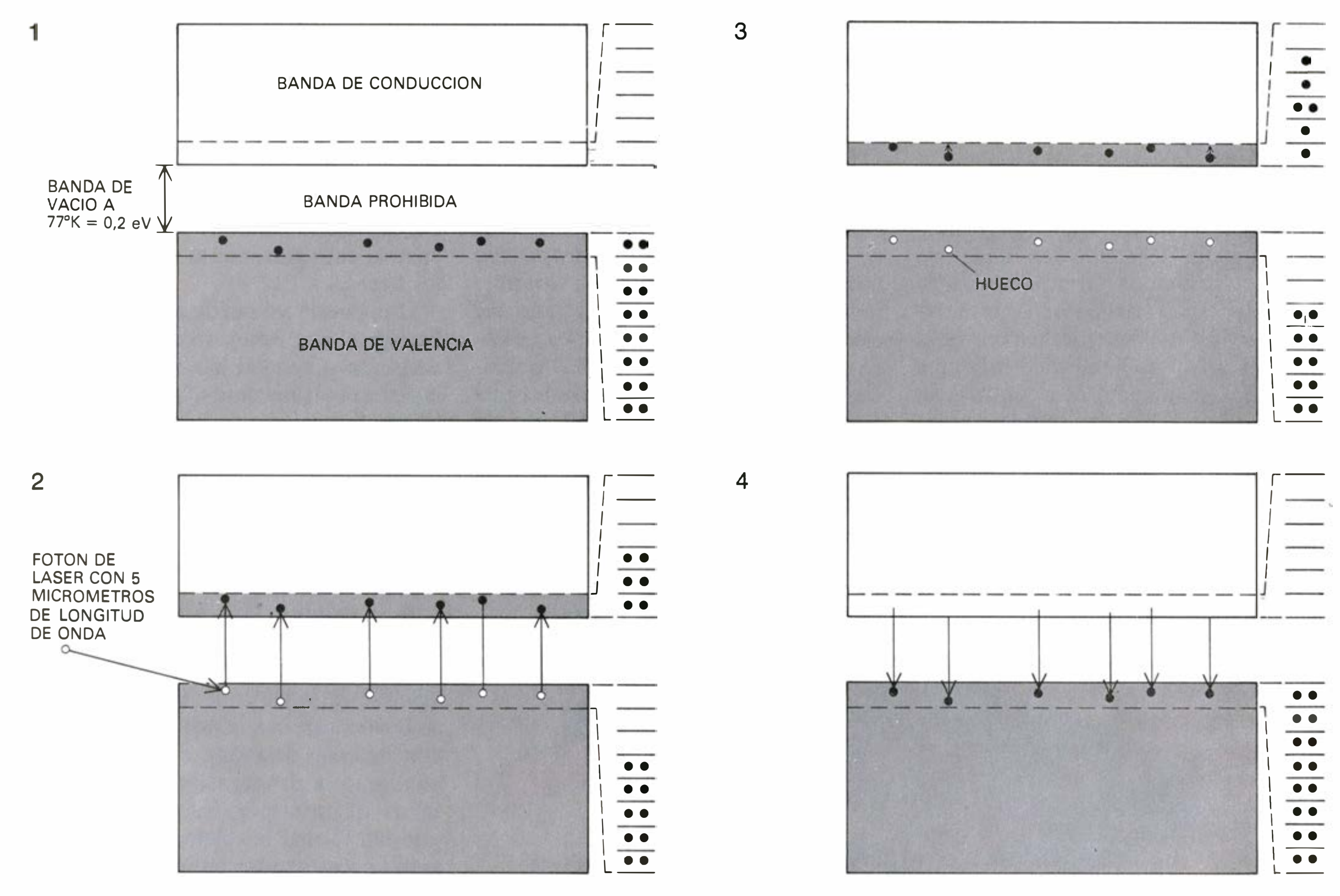
La energía requerida por un electrón para cruzar el intervalo prohibido varía con la temperatura. Buena parte de nuestro trabajo con antimoniuro de indio se ha realizado a 77 grados Kelvin. A esa temperatura, la energía necesaria para cruzar el intervalo prohibido es de

unos 0,2 electronvolt. (Un electronvolt, abreviadamente eV, es la energía que gana un electrón cuando se le acelera a través de un potencial de 1 volt.) Muchos otros semiconductores tienen un intervalo prohibido más amplio. La no linealidad del índice de refracción aumenta en razón inversa del cuadrado de la energía del intervalo prohibido. Debido a su estrecho intervalo prohibido, el antimoniuro de indio tiene una no linealidad entre 100 y 1000 veces mayor que el arseniuro de galio, otro semiconductor utilizado en dispositivos ópticos biestables. Aunque es conveniente que el intervalo prohibido sea estrecho, cuando lo es demasiado la operación de selección óptica resulta inviable.

El electrón adquiere la energía necesaria para su promoción a la banda de conducción absorbiendo la radiación incidente. De ahí que ahora convenga adoptar la descripción de la luz como un flujo de partículas. La partícula, o "cuanto", de luz, es el fotón. La energía de un fotón es inversamente proporcional a su correspondiente longitud de onda. Cuando un fotón es absorbido por el cristal, su energía puede sumarse a la de un electrón. Si el fotón transporta la cantidad necesaria de energía, el electrón pasa a la banda de conducción. La longitud de onda equivalente a 0,2 eV es de unos 5 micrometros, la cual está dentro de la zona infrarroja del espectro y dentro de la gama de nuestro láser de monóxido de carbono.

Ajustando el láser a la longitud de onda de 5 micrometros, conseguiremos que muchos electrones pasen a la banda de conducción. En cuanto éstos alcanzan los niveles más bajos de la banda de conducción, la energía térmica de la estructura cristalina provoca un esparcimiento de los electrones a otros niveles de energía próximos. Resulta de todo ello una distribución irregular en el tramo inferior de la banda de conducción.

Las causas que modifican el índice de refracción del antimoniuro de indio son la excitación y consiguiente difusión de los electrones. En la descripción mecanicocuántica de la interacción entre luz y materia, la refracción y la absorción están íntimamente ligadas. Una implica la otra: si un haz de radiación resulta absorbido, también es refractado y viceversa, aunque los dos fenómenos puedan afectar a radiaciones de diferentes longitudes de onda dentro del haz. Previendo la absorción, antes de que tenga lugar, podemos también prever la refracción, antes de que se produzca, y de esta manera alterar el índice de refracción. ¿Cómo prever la absorción? Existe un límite en el número de electrones que pueden ser excitados para alcanzar los niveles más bajos de la banda de conducción. Cuando se alcanza dicho límite ya no se pueden seguir



8. INDICE DE REFRACCION del antimoniuro de indio. Varía en función de la intensidad de la radiación que lo atraviesa. Se trata del semiconductor empleado por los autores como material no lineal del transistor óptico. En un cristal semiconductor, los niveles de energía permitidos para los electrones se disponen en bandas. Dentro de cada banda hay muchos niveles próximos entre sí, cada uno de los cuales puede ser ocupado por dos electrones; no así los niveles energéticos de los espacios vacíos entre estas bandas, que están prohibidos. Se dibujan aquí dos bandas: a baja temperatura, la banda inferior (llamada la banda de valencia) está ocupada y, vacía, la banda superior, o banda de conducción (*1*). Para ir desde la banda de valencia hasta la de conducción en el proceso denominado excitación el electrón ha de adquirir la energía equivalente a la representada por la banda prohibida; en el antimoniuro de indio esta energía es de unos 2 electronvolt (eV). Semejante energía puede lograrse absorbiendo un fotón con una longitud de onda de 5 micrometros, ubicada en el infrarrojo. Los autores emplearon un láser de CO con una longitud de onda capaz de generar tales fotones (*2*). Cuando un electrón absorbe un fotón y asciende hasta la banda de conducción, deja atrás, en la banda de valencia, un hueco que actúa cual una partícula de carga positiva. La energía térmica y de vibración desplazan rápidamente los electrones entre los niveles más bajos de la banda de conducción (*3*). Cuando estos niveles están ocupados, no pueden absorberse más fotones de 5 micrometros de longitud de onda. El bloqueo de la absorción permite un cambio simultáneo en el índice de refracción. Al desconectar el haz incidente, los electrones vuelven a la banda de valencia y se recombinan con los huecos, que así desaparecen (*4*).

absorbiendo nuevos fotones de la energía precisa para producir estas excitaciones: la absorción de dichos fotones se dice que está saturada. La saturación de la absorción impide la refracción asociada y el índice de refracción cambia. Muy poco después que se apague el haz láser, los electrones excitados vuelven a la banda de valencia y se recombinan con los huecos. Entonces torna a ser posible la absorción y el índice de refracción alcanza de nuevo su valor original.

Puertas lógicas ópticas

Hemos hecho uso de las propiedades ópticas del antimoniuro de indio, únicas, que resultan de este mecanismo para construir los elementos de un computador óptico. El ciclo de histéresis, descrito más arriba, tiene una aplicación obvia como elemento de memoria. El estado estable de alta transmisión del interferómetro puede representar un valor lógico o aritmético; el estado estable de baja transmisión, el otro valor. Cada estado puede mantenerse indefinidamente por un haz de intensidad media.

No necesitamos la histéresis para utilizar un interruptor óptico como procesador de la información. Modificando levemente la estructura óptica, se reduce, o se elimina del todo, el lazo de histéresis. El resultado será un valor único en la curva de transmisión, en la que cada nivel de intensidad incidente se asocia con un sólo nivel de intensidad transmitida. La forma de esta curva univaluada es susceptible de manipulación. En una de las formas más comunes la transmisión es baja y casi constante a baja intensidad incidente; luego, al llegar a cierto umbral, la transmisión asciende bruscamente y alcanza un nivel alto que permanece de nuevo casi constante, aunque la intensidad incidente aumentara. La curva se asemeja a la que describe la corriente procedente del colector de un transistor electrónico; es la base del transistor óptico. Nosotros damos el nombre de transfasor a este dispositivo porque su operatividad está fundamentada sobre el control de la fase de la luz en su interior.

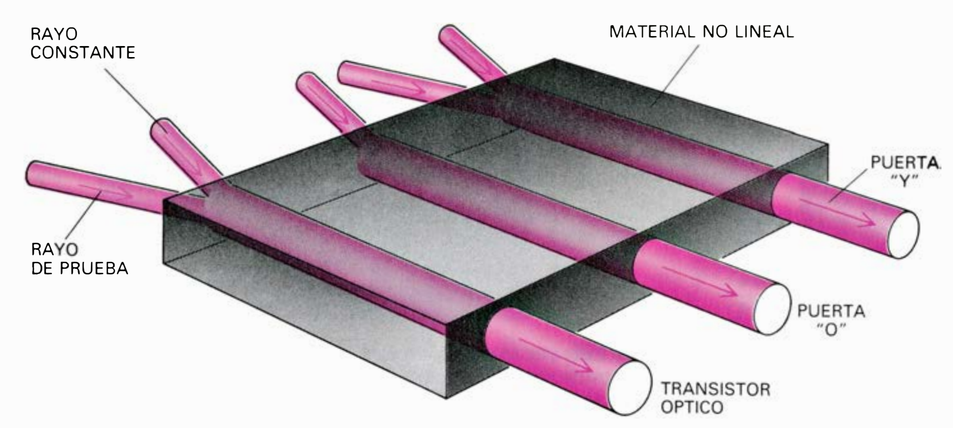
En el transfasor, dos haces láser ajustados se enfocan hacia un mismo punto en la cara frontal del cristal de un material no lineal, antimoniuro de indio por ejemplo. El haz "constante" tiene una intensidad grande e invariante. El haz de "prueba" tiene una intensidad mucho menor y puede ser modulado. La potencia del haz constante se determina de manera que proporcione una intensidad transmitida justamente por debajo de la zona abrupta de la curva de transmisión. La intensidad del haz de prueba deberá ser tal que, cuando se sume a la intensidad del haz constante, se supere dicha zona y pueda alcanzarse el máximo de transmisión.

Al ser la curva de transmisión del transfasor muy abrupta, una ligera modificación del haz de prueba controla un elevado aumento en la intensidad transmitida del dispositivo. Ampliemos, pues, su analogía con el transistor electrónico. El haz constante es análogo a la corriente constante de "polarización" que fluye desde el emisor hasta el colector del transistor. El haz de prueba, por su parte, recuerda la corriente menor que fluye desde la base hasta el colector. De la misma forma que un pequeño cambio en la corriente de la base posibilita al transistor para conducir una corriente mucho mayor al colector, un pequeño cambio en el haz de prueba capacita al transfasor para transmitir una mayor fracción de la radiación incidente.

Al igual que el transistor, el transfasor puede ser conmutado entre dos estados diferenciados. La selección de uno u otro es extremadamente rápida. Aunque nosotros no hemos logrado medir directamente el tiempo de conmutación, las pruebas indirectas nos sugieren que es de unos pocos picosegundos.

Una vez que se ha construido el transistor óptico, el montaje del prototipo de una puerta lógica no encierra ninguna dificultad. Según los haces aplicados podría servir un transfasor único, como puerta Y o como "O". Si se seleccionan los dos haces incidentes de forma que ninguno de ambos baste por sí solo para alterar el dispositivo, y juntos tengan la suficiente energía para hacerlo, el resultado es una puerta Y óptica. Unicamente cuando ambos haces incidentes están presentes la transmisión será elevada. Si se ha seleccionado el haz incidente de manera que por sí solo sea capaz de conmutar el dispositivo, el resultado será una puerta "O" óptica. La intensidad transmitida será elevada cuando esté presente al menos uno de los haces.

Una puerta NO puede construirse utilizando el haz reflejado como terminal de salida. Como el haz reflejado es el inverso del transmitido, aumentando la intensidad incidente bajará el rendimiento en la salida. Inversamente, si disminuimos el haz incidente, aumentará el rendimiento en la salida.



9. ES EL PROCESADO EN PARALELO una propiedad del interruptor óptico que nos llevará a nuevos diseños de los computadores. Podrían enfocar múltiples haces láser que permanecerían separados en su paso por un material no lineal. Cada trayecto de radiación podría servir para realizar tareas diversas; podrían ser distintas las operaciones llevadas a cabo con varios haces. Si hubiese tres pistas, por ejemplo, una valdría de puerta Y, otra de O y la tercera de transistor. Con tres haces incidentes, procedentes de un mismo haz que ha sido fraccionado con anterioridad, pueden verificarse tres operaciones de golpe. Tal potencialidad requeriría una forma nueva de procesamiento de información; en los ordenadores electrónicos sobre una misma señal de partida sólo puede hacerse una operación. (Diseño de Walken Graphics.)

Hacia la computación fotónica

Para construir un computador óptico, deberían montarse e interconectarse muchos elementos de los circuitos ópticos. A lo largo del último decenio se llevó a cabo un considerable trabajo con películas delgadas, a cuyo través se consiguió la propagación de los haces láser. En principio, las guías de onda podrían formar los elementos de conexión en los circuitos ópticos integrados, si bien no se ha abordado aún la fabricación de dichos circuitos.

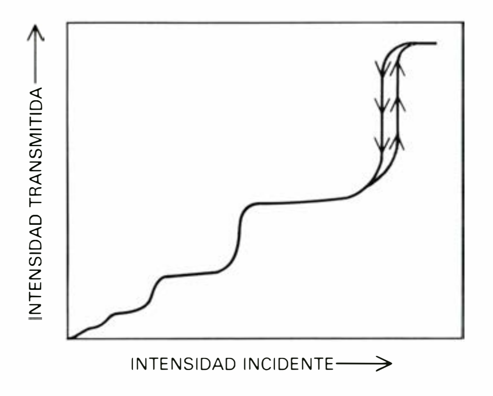
Una aplicación de esta nueva tecnología óptica podría ser la construcción de computadores con estructura u organización lógica similar a la de las máquinas electrónicas convencionales. El interés principal de un computador con estructura convencional, pero dotada de componentes ópticos, podría ser su

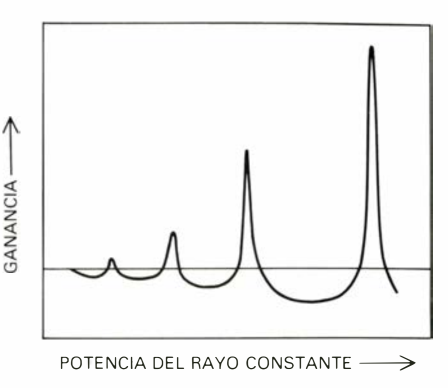
gran velocidad. Las puertas lógicas ópticas podrían llevar a cabo un billón de operaciones lógicas por segundo, en comparación con los mil millones para los dispositivos selectores electrónicos.

Los conmutadores ópticos tienen, en otro orden, propiedades que podrían aplicarse para nuevos diseños de computadores. Mientras que en un transistor varias corrientes que circulen simultáneamente llegan inevitablemente a mezclarse, a través de un interferómetro podrían pasar simultáneamente múltiples haces y permanecer separados. De esta manera bastaría un cristal para realizar varias operaciones de conmutación distintas. Podríamos disponer, por ejemplo, de cinco pistas láser a través del cristal, cada una de las cuales fuese el lugar de una operación. Seleccionando adecuadamente la intensidad del haz, cada operación sería diferente. Dos pistas funcionarían como puertas Y, otras dos como "o" y una quinta operaría de la misma forma que en el transistor básico.

Si los haces modulados para cada una de las cinco pistas se enviaran desde una señal única, previamente dividida en cinco partes, la señal original podría estar sometida a cinco operaciones lógicas a la vez. Después, las señales resultantes podrían procesarse, cada una por separado, o bien juntas. Esta nueva forma de operar requeriría circuitos computadores muy diferentes de los que posee una máquina electrónica.

El interruptor óptico posibilitaría, además, otras modificaciones, lógicas y de diseño. Arriba mencionamos una posibilidad: introducir dispositivos que tengan más de los dos estados lógicos. Algunos de los cristales con los que hemos estado trabajando tienen capacidad para alcanzar sucesivamente mayores niveles de intensidad transmitida, con sucesivos incrementos en la intensidad del haz incidente. Los niveles de intensidad transmitida pueden representar los estados de un sistema con múltiples valores lógicos.





10. BIESTABILIDAD MULTIPLE, propiedad del interruptor óptico que podría conducir al desarrollo de una nueva lógica para computadores. Para un interferómetro, hay muchas longitudes de camino óptico iguales a un múltiplo entero de semilongitudes de onda de radiación. Cada longitud óptica corresponde a un máximo en la función de Airy, alcanzando así máxima transmisión. Aumentando la intensidad del haz incidente conseguiremos sucesivos máximos, al ir modificándose la longitud óptica del cristal no lineal de la cavidad. El diagrama de la intensidad transmitida en función de la intensidad incidente muestra varios picos muy abruptos, cada uno de los cuales corresponde a un máximo en la curva de Airy (*izquierda*). Si incrementamos la intensidad del haz constante en el transistor óptico, las sucesivas activaciones del haz de prueba permiten, con una pequeña potencia, alcanzar las sucesivas elevaciones verticales en la curva de intensidad transmitida. Desde el momento en que los ascensos verticales de la curva de la izquierda se hacen progresivamente más grandes, el incremento o ampliación de la energía de salida por operación del haz de prueba aumenta en cada operación de selección (*derecha*). Los niveles de transmisión podrían asociarse a un nuevo sistema lógico que reemplazara el sistema lógico binario.

Procesamiento óptico de datos

Los sistemas de conmutación ópticos no han de concebirse necesariamente como meros sustitutos más rápidos de los dispositivos electrónicos. Antes bien, los mayores beneficios de los interruptores ópticos derivarán de sus nuevas aplicaciones. Las fibras ópticas, gracias a su prodigiosa capacidad de transportar información, se están generalizando en el campo de las comunicaciones, incluidas las propias comunicaciones entre computadores. El interruptor óptico constituye el candidato natural para mediar entre los sistemas electrónicos y los ópticos. Por otra parte, si la computación se hace ópticamente al principio, las fibras ópticas podrían emplearse como lazos directos entre sistemas de computación.

Según nuestras comprobaciones, los interruptores ópticos pueden prepararse en forma de delgados cristales. La tecnología actual del crecimiento de cristales o de películas delgadas posibilita la fabricación de grandes láminas delgadas. De este modo, si una imagen óptica se proyectara directamente a través de una lámina grande, en la que cada pequeña superficie sirviera de conmutador óptico, la transmisión desde los interruptores serviría de registro digital de la imagen. Podrían efectuarse múltiples formas de procesado y ampliación de las imágenes mientras se llevara a cabo la grabación.

Hemos manifestado hasta aquí nuestro entusiasmo por el potencial de estos dispositivos de conmutación óptica. Bueno será recordar algunas dificultades que se presentan cuando se pasa del ensayo experimental a la comercialización del producto. En nuestras investigaciones hemos tenido que trabajar a bajas temperaturas. Si se pudiese operar a una temperatura ambiente o casi ambiente el interruptor óptico resultaría obviamente más práctico. El intervalo prohibido de energía del antimoniuro de indio es pequeño y disminuye al aumentar la temperatura. Empleando un láser de monóxido de carbono con una longitud de onda de 10 micrometros, conseguimos un interruptor óptico a temperatura ambiente. Y es indudable que resultará posible mejorar la velocidad y eficacia del interruptor con otros tipos de láser y de materiales no lineales. Conocer el material adecuado constituye el punto central de toda construcción de máquinas ópticas de computación. En teoría, el mejor sería un interruptor óptico que consumiera la menor potencia posible. Se llama potencia de conmutación al producto de la intensidad de la luz incidente en el cristal por el área de la cara del cristal. Reduciendo el tamaño del dispositivo, rebajaremos también su potencia de conmutación. Pero ei dispositivo no puede reducirse indefinidamente: existe un límite en la forma en que se puede enfocar el haz láser. Resumiendo, si el tamaño del dispositivo disminuye, el sobrecalentamiento se convierte en un problema serio.

Ante la pretensión de una potencia de interrupción baja, recuérdese que el tiempo de conmutación puede acortarse, dentro de ciertos límites, aportando más potencia. La magnitud del dispositivo, la velocidad de conmutación y su potencia deben considerarse siempre en función unos de otros, en función de las propiedades ópticas y térmicas del medio no lineal y en función del interruptor. La selección del medio correcto o el descubrimiento de nuevos materiales no lineales es, probablemente, la tarea más importante a abordar en el desarrollo del interruptor óptico. Sin desdeñar tan formidables dificultades, creemos que el computador óptico constituye un objetivo realista y fascinante.

Ranas venenosas

En Colombia, los cazadores indios emponzoñan los dardos de las cerbatanas con alcaloides tóxicos que segregan unas pequeñas ranas. Tanto los alcaloides como la biología evolutiva de esos anfibios plantean interesantes cuestiones

Charles W. Myers y John W. Daly

Los organismos ponzoñosos son tan numerosos que químicos y farmacólogos se hallan todavía a muchos años de distancia de poder analizar todas las moléculas nocivas que se han encontrado en varios microorganismos, plantas y animales. La investigación básica en este tema interesa a campos tan extensos como la ecología y la biología evolutiva; puede conducir asimismo a nuevos medicamentos y otras sustancias útiles. La riqueza de vida de las regiones tropicales proporciona ejemplos de particular interés. Uno, perteneciente a los trópicos del Nuevo Mundo, es el género vegetal *Strychnos*, fuente del veneno estricnina y de los alcaloides del curare. Los cazadores indios utilizan curare para envenenar flechas y dardos de cerbatana, y los anestesistas lo administran para relajar los músculos. Otro ejemplo procede de una familia de ranas, algunas de cuyas especies segregan uno de los venenos animales más poderosos que se conocen. También esta sustancia se aprovecha para envenenar dardos y es de aplicación en investigación biomédica. Hay especies de esa misma familia que segregan otras toxinas, y las mismas ranas centran la atención por aspectos muy peculiares de su biología.

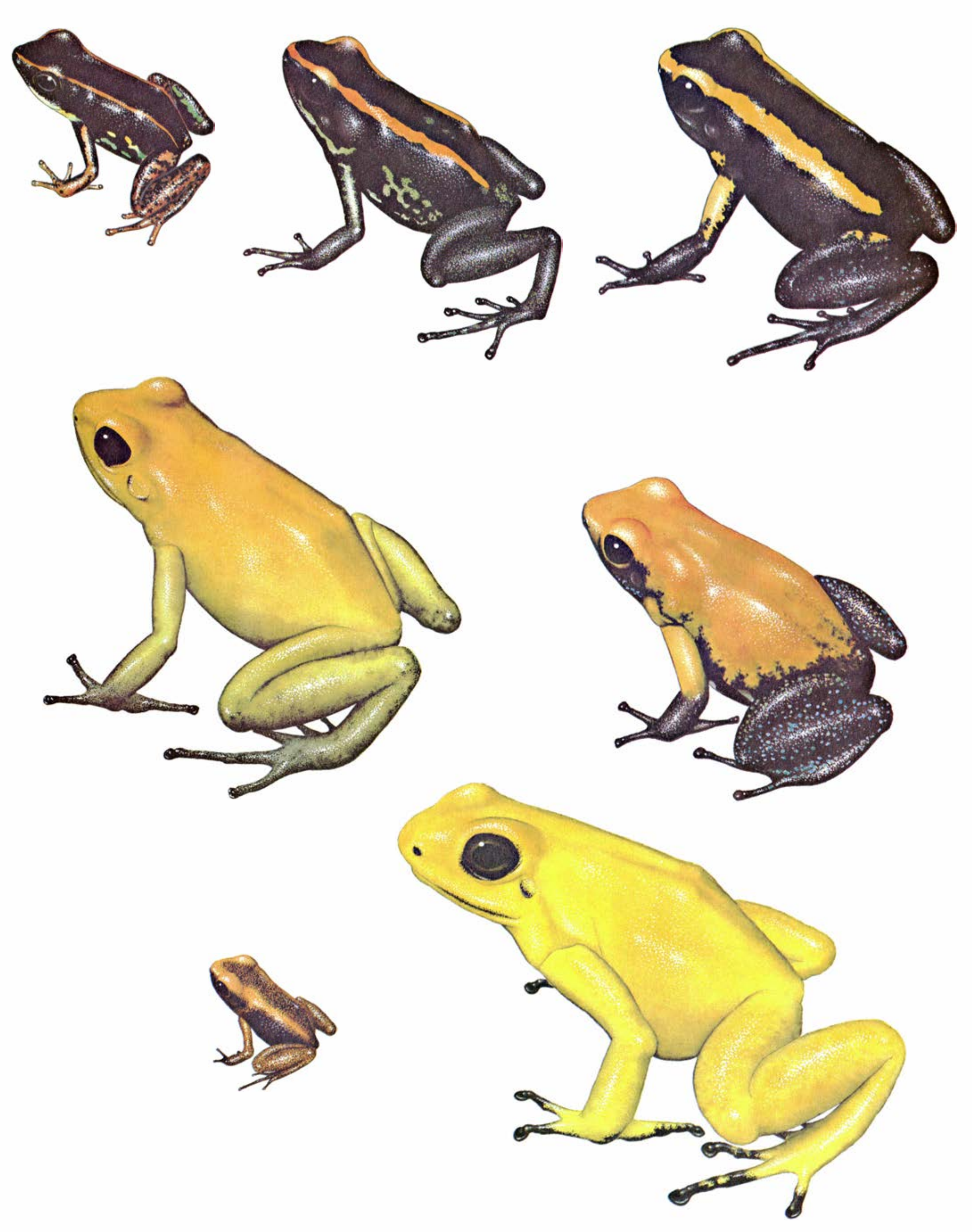
Las ranas venenosas y sus parientes no tóxicas pertenecen a la familia Dendrobátidos, acantonada en Sudamérica y en el sur de Centroamérica. Ocupan una gran variedad de hábitats. Algunas especies medran a lo largo de los ríos; otras viven lejos del agua, sobre o cerca del suelo del bosque lluvioso de las llanuras o montano. Unas cuantas especies forestales pasan la mayor parte de su vida en lo alto de los árboles. En el otro extremo, las hay que viven en terreno abierto y seco, donde encuentran la humedad suficiente sobre el suelo umbrío, bajo la vegetación baja. A pesar de esta diversidad ecológica, las ranas dendrobátidas comparten un estilo de vida que las distingue de las demás. Sólo son activas durante el día, y ponen sus huevos en tierra, en lugares húmedos. Cuidan de los huevos hasta que hacen eclosión; momento a partir del cual una rana "nodriza" transporta a los renacuajos, literalmente pegados a su dorso, hasta un ambiente acuático adecuado.

La familia Dendrobátidos comprende más de 100 especies, que generalmente se dividen entre cuatro géneros. Son ranas pequeñas, que oscilan en longitud, cabeza y cuerpo, desde poco más de un centímetro hasta unos cinco centímetros. Mientras que en la mayoría de ranas el macho es menor que la hembra, en muchas poblaciones de dendrobátidos macho y hembra son de tamaño similar. Los machos grandes presentan un elevado grado de territorialismo y agresividad; e invierten mucho tiempo y energía en advertir de su territorio y en atraer a una pareja mediante llamadas prolongadas, así como en desafíos y escaramuzas con otros machos. Muchos machos territoriales son pequeños seres ruidosos; sus gorjeos, píos y trinos parecidos a los de los insectos contribuyen de forma apreciable a los sonidos diurnos de algunos bosques tropicales. El espectáculo de dos ranitas diminutas luchando sobre una hoja, lanzándose mutuamente vigorosos chirridos puede ser divertido para un observador humano, pero la eficacia sexual y el éxito reproductor son asuntos serios desde el punto de vista evolutivo. En algunas especies de dendrobátidos la hembra también es agresiva y defiende el territorio.

El tamaño relativamente grande de los machos de dendrobátidos determina asimismo un comportamiento copulador insólito. En la mayoría de anuros, el cortejo concluye al montar el macho a la hembra, más grande; el macho agarra a su compañera por detrás de las patas anteriores de ésta o bien por delante de sus patas posteriores. Los machos de los dendrobátidos sujetan a la hembra por la cabeza, no alrededor del tronco, y con la parte posterior de los dedos hacen presión debajo de la mejilla de la hembra. Esta posición anterior asegura mejor la fecundación de los huevos que va expeliendo la hembra; si un macho de tamaño similar se situara más hacia atrás, su esperma caería al suelo. En algunas especies de dendrobátidos no hay abrazo en el ritual de cortejo, y los huevos se fecundan en el curso de maniobras bastante complicadas que realizan tanto la hembra como el macho.

Según las especies, los huevos de los Dendrobátidos se ponen en la hojarasca, en grietas que quedan bajo las rocas, sobre hojas sombreadas a cierta altura del suelo o bien en plantas propias del bosque tropical, como bromeliáceas y aráceas. El tamaño de la puesta de huevos es pequeño en comparación con los cientos o incluso miles de huevos que ponen muchas de las ranas que se reproducen en el agua; puede ser de sólo uno o dos huevos en los dendrobátidos más pequeños, y raramente tiene más de 30 o 40. En muchas poblaciones el reducido tamaño de la puesta se compensa en parte con la reproducción continua a lo largo de todo el año. Ciertas ranas pueden reproducirse cada mes. La observación de una pareja de ranitas de la especie *Dendrobates tricolor*, mantenida en el Museo Americano de Historia Natural durante más de dos años, apoya con datos estadísticos este punto. Las ranas se apareaban de dos a cuatro veces al mes; por término medio, había puestas cada 10 días, de 12 a 30 huevos cada una. Incluso con este esfuerzo prodigioso, la hembra no alcanzó a poner más que unos 600 huevos al año.

Una vez depositados, los huevos quedan bajo la protección de uno de los progenitores, cuando no se abandonan; reciben breves visitas periódicas, en las que la rana los humedece a ve-

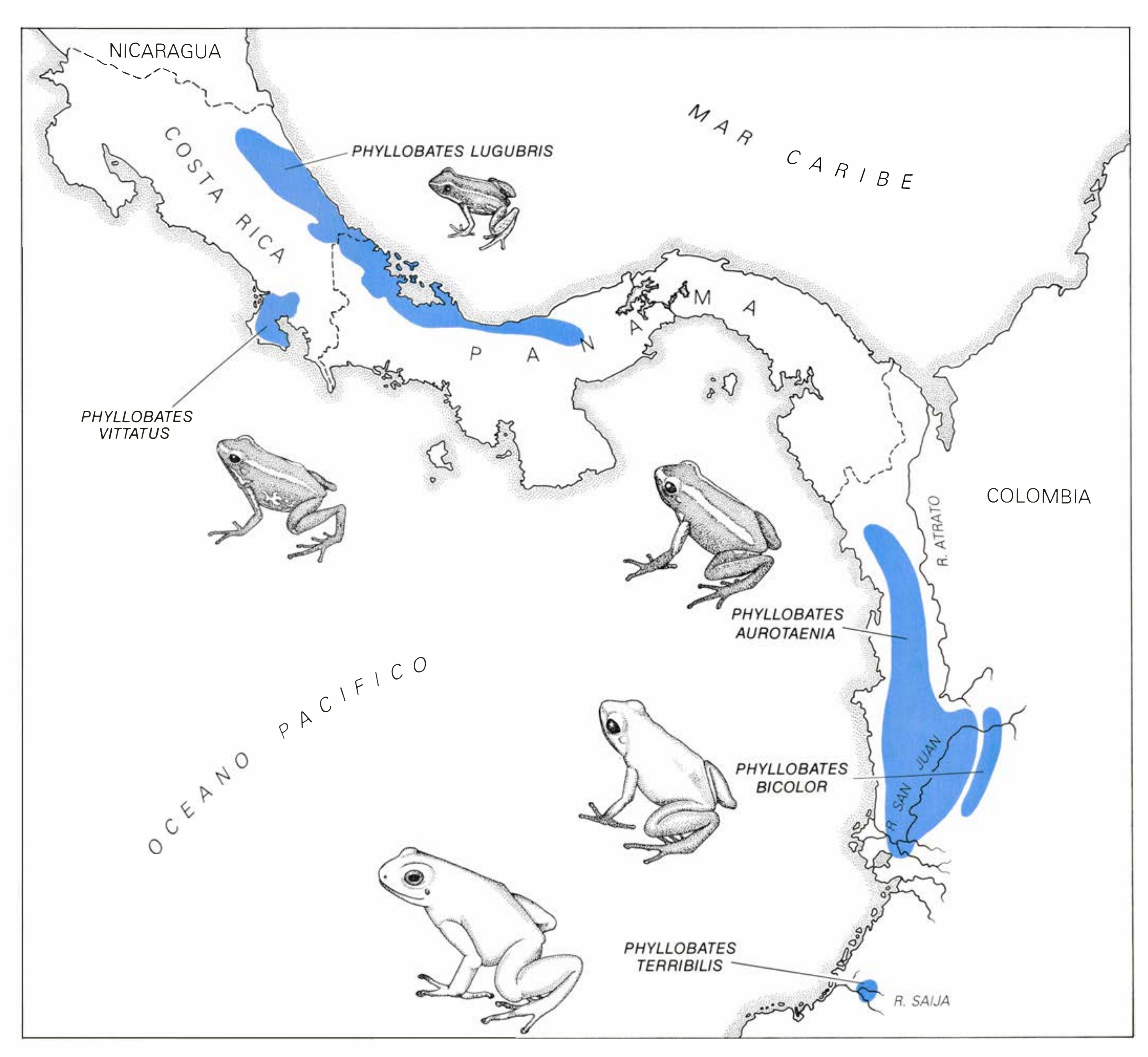


1. RANAS PONZOÑOSAS del género *Phyllobates*, reproducidas a 1,7 veces su tamaño real en estas pinturas de David M. Dennis. Los dos primeros ejemplares de arriba, empezando por la izquierda, son *P. lugubris* y *P. vittatus*, especies centroamericanas que segregan toxinas a través de glándulas de su piel. La rana listada de la parte superior derecha es *P. aurotaenia*, una de las tres especies colombianas utilizadas para emponzoñar dardos de cerbatana. En la fila del medio hay dos ejemplares de *P. bicolor*, la segunda de las tres especies colombianas. Muestran aquí dos situaciones extremas de la variabilidad en tamaño y color; las patas, de un color distinto al del cuerpo, van del verde pálido al negro. Abajo aparece la tercera especie colombiana, la más venenosa, *P. terribilis*. El inmaturo de esta especie (*izquierda*) es listado, como sus parientes de arriba, pero sus listas resultan obliteradas por la extensión de pigmento brillante. El adulto de *P. terribilis* (*derecha*) puede ser amarillo, anaranjado o verde pálido, pero por lo general se halla uniformemente coloreado, dorsal y ventralmente. *P. terribilis* es una rana más atrevida que las demás y es, al menos, veinte veces más tóxica. Puede ser peligroso tocarla.

ces, aparentemente con agua de su vejiga. La rana nodriza hace colear ocasionalmente sus cuartos traseros en la masa de huevos, acción que, en última instancia, facilita la eclosión y ayuda a los renacuajos a trepar, serpenteando, por el dorso del progenitor. Allí se adhieren a una mancha de mucus segregada por glándulas de la piel de la rana. El vientre plano o ligeramente cóncavo característico de los renacuajos de los Dendrobátidos facilita la adhesión. En algunas especies, la fijación es sólo superficial; en otras, el mucus adquiere la consistencia de la cola, y los renacuajos quedan firmemente sujetos al dorso del adulto. También según las especies, los renacuajos permanecen fijados a la rana nodriza durante un intervalo que va de unas pocas horas a más de una semana. En este período, los renacuajos se desarrollan absorbiendo reservas de vitelo.

Llegado su momento, la rana nodriza se acerca hasta una charca o curso de agua; allí, un remojo suficiente libera la unión mucosa y los renacuajos nadan libremente. Muchos dendrobátidos llevan sus renacuajos a arroyos. Otras especies los sueltan en las pequeñas charcas de agua de lluvia que se acumulan en las bromeliáceas o entre la hoja y el pecíolo de algunas otras plantas tropicales. Las ranas nodriza de una especie, *Dendrobates auratus*, suelen liberar a sus renacuajos en el agua retenida en los agujeros de troncos de árboles. Los renacuajos de *D. auratus* tienden a recurrir al canibalismo en este ambiente pobre en nutrientes, y puede que sólo uno de ellos alcance la madurez.

El agua de lluvia acumulada en las pequeñas bromeliáceas y en las axilas de las hojas parece retener asimismo pocos nutrientes, circunstancia que ayuda a explicar una sorprendente observación protagonizada por Peter Weygoldt, de la Universidad alemana de Friburgo. Trabajando con ejemplares de laboratorio de *Dendrobates pumilio*, una especie que suelta sus renacuajos en tales lugares, Weygoldt descubrió que las hembras visitaban regularmente a sus renacuajos de vida libre para depositar en el agua huevos no fecundados; éstos servían luego de alimento a los renacuajos.



2. DISTRIBUCION IRREGULAR de las cinco especies del género de dendrobátidos *Phyllobates*. Las tres especies que habitan en Colombia occidental son, que se sepa, las únicas ranas utilizadas por los cazadores indios para emponzoñar dardos de cerbatana. Las dos especies de Costa Rica y Panamá segregan pequeñas cantidades de las mismas toxinas que sus parientes colombianas, pero no viven en las regiones donde se emplean las cerbatanas.

Los biólogos intentan a veces determinar el "costo" relativo del cuidado paterno que imparten los machos o las hembras, porque tal conducta puede imponer limitaciones a la reproducción o puede exponer al progenitor a la depredación. Son éstos factores que hay que sopesar frente al beneficio de asegurar el éxito a uno o unos pocos grupos de huevos o al esperma de un determinado padre. En las ranas dendrobátidas, el costo reproductor puede ser superior en especies (*D. pumilio* entre ellas) en las que las ranas nodriza son por lo general hembras, que en aquellas en las que la crianza corresponde a los machos. El macho de la pareja de *D. tricolor* del Museo Americano podía cuidar a un tiempo varias puestas de huevos. Además, vocalizaba con frecuencia y presumiblemente se hubiera apareado con otras hembras de haberlas tenido a su alcance. Comía incluso mientras llevaba en su dorso una carga completa de renacuajos, que liberó al agua, pocas horas tras la eclosión.

Se ha supuesto que el sexo de la rana nodriza (macho en la mayoría de especies y hembra en algunas) es una característica específica. Sin embargo, en unas pocas especies se ha visto que ambos sexos acarrean los renacuajos, algo difícil de explicar. ¿Cómo puede ser tan variable un rasgo de comportamiento tan reciente? Experiencias con la pareja de *D. tricolor* sugieren que la respuesta puede radicar en el resultado de una competencia sexual por el papel de nodriza. La hembra de la pareja podía y deseaba ser el progenitor responsable, pero la agresividad territorial del macho le impedía desempeñar ese papel. La hembra quedaba al cuidado de la puesta de huevos durante la primera hora, aproximadamente; después, no se le permitían ulteriores derechos de visita. Sin embargo, si se apartaba el macho después de la cría, la hembra cuidaba de los huevos y hasta acarreaba con los renacuajos. Si el macho no se apartaba hasta unos días después, la hembra no mostraba reparo alguno en comerse los huevos o los renacuajos, como si pertenecieran a otra hembra. Tal canibalismo puede constituir un mecanismo de regulación de la densidad de población en algunas especies de dendrobátidos.

Queda todavía mucho por descubrir sobre la historia natural de los Dendrobátidos, sus relaciones evolutivas y su clasificación. Las aproximadamente 130 especies suelen agruparse en cuatro géneros: *Atopophrynus, Colostethus, Dendrobates* y *Phyllobates*. *Atopophrynus* tiene una única especie. *Colostethus*, más de 70, principalmente de color pardo; con unas pocas excepciones, estas especies no son tóxicas. La mayoría de las aproximadamente 50 especies de *Dendrobates* se distinguen por una brillante coloración advertidora. Los vivos colores de estas ranas advierten de la presencia de secreciones dérmicas ponzoñosas, o al menos desagradables al paladar, que son efectivas a la hora de disuadir a muchos depredadores potenciales. Sin embargo, los dendrobátidos tóxicos no son completamente inmunes a la depredación; por ejemplo, son presa fácil de grandes arañas y algunas serpientes. Las secreciones de las ranas son liberadas por glándulas microscópicas de la piel en momentos de tensión; un depredador que haya aprehendido a una rana con la boca experimentará sensaciones de quemazón, entumecimiento o mal sabor que probablemente le hagan soltar la presa.

Los dendrobátidos más tóxicos son tres especies de *Phyllobates* del oeste de los Andes, en la vertiente pacífica de Colombia. Estas ranas segregan un veneno mucho más potente que el curare. Ampliamente utilizado éste para las flechas y los dardos de cerbatana al este de los Andes, las toxinas de estos dendrobátidos empozoñan los dardos en Colombia occidental. Posiblemente el relato más antiguo de su utilización sea el del capitán Charles Stuart Cochrane, que exploró Colombia en 1823-24, disfrutando de un permiso de la Armada británica.

Mientras cruzaba a pie los Andes occidentales, se fijó en unas ranas "llamadas *ranas de veneno* [por los españoles], de unos siete centímetros de longitud, de color amarillo en el dorso, con ojos negros muy grandes". "Los que utilizan [su] veneno capturan las ranas en los bosques y las encierran en una caña hueca, donde les dan de co-



3. CUIDADO DE LAS CRIAS por parte del progenitor. Macho o hembra vigilan los huevos. Aquí, un macho de *Dendrobates silverstonei* monta guardia junto a la puesta depositada sobre una hoja del suelo.

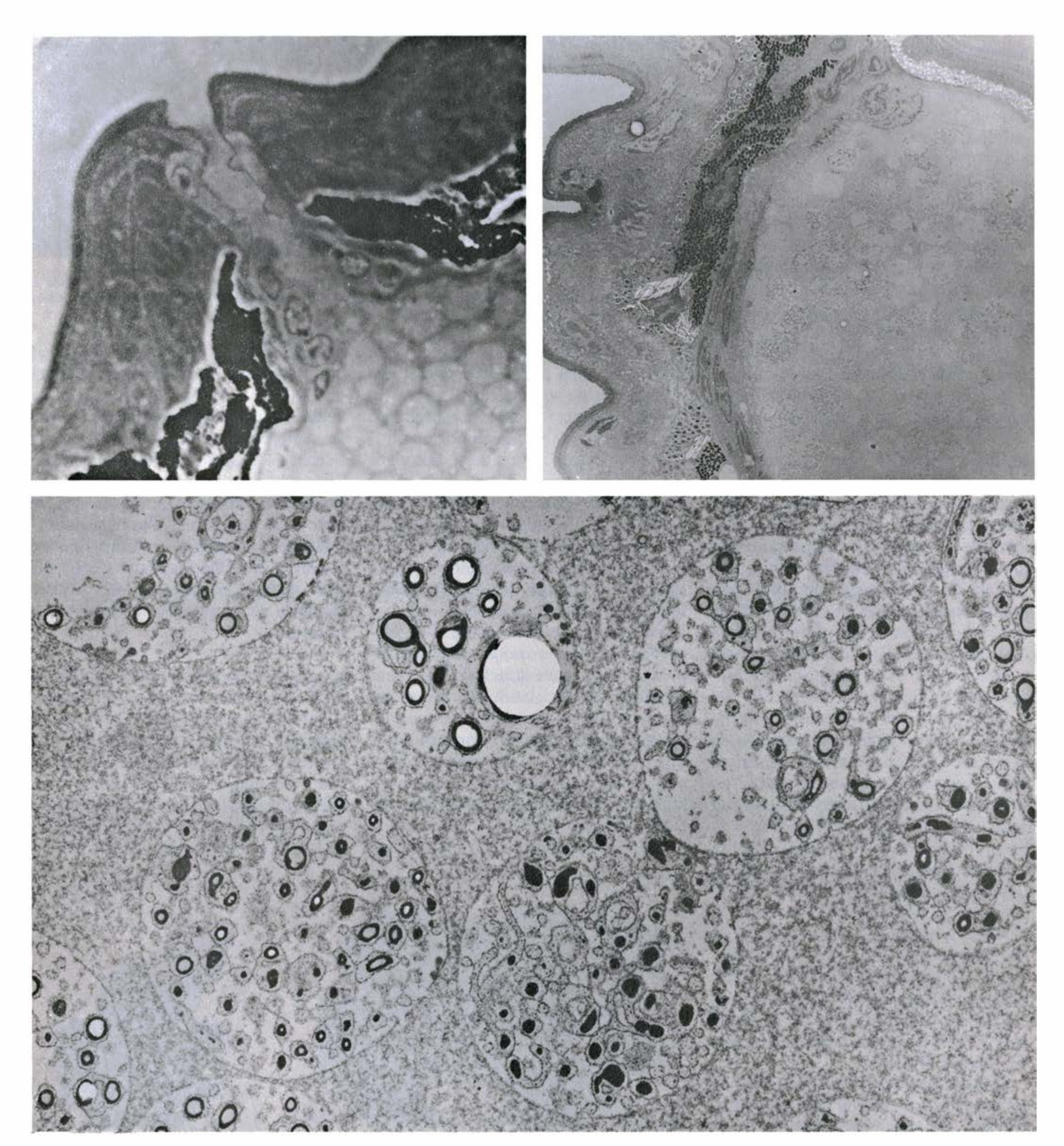




4. RANAS "NODRIZA", que transportan a sus renacuajos a una masa de agua adecuada después de la eclosión. Una hembra de la especie *Colostethus inguinalis*, de Panamá (*izquierda*), aparece aquí con una carga de 27 renacuajos. Un macho de la especie peruana *Dendrobates reticulatus* (*derecha*) transporta sólo uno. La cría permanece pegada de esta manera a su progenitor horas, o hasta más de una semana.

mer regularmente hasta que necesitan el veneno, momento en el que toman uno de los infortunados reptiles y le insertan un fragmento puntiagudo de madera por la garganta, hasta que sale por una de sus patas. La tortura hace que la pobre rana sude mucho, especialmente en el dorso, que se cubre de una espuma blanca: ésta es el veneno más poderoso que suministra, y en ella mojan o hacen rodar las puntas de sus flechas, que conservarán su poder destructor durante un año. Después, debajo de esta sustancia blanca aparece un aceite amarillo, que rebañan cuidadosamente, y que retiene su mortífera influencia de cuatro a seis meses, según la bondad (como dicen ellos) de la rana. Por este procedimiento, de una rana se obtiene veneno suficiente para unas cincuenta flechas".

Parece que *Viajes por Colombia* de Cochrane no tuvo muchos lectores. En todo caso, unos cuantos observadores



5. GLANDULAS DERMICAS "GRANULARES" que almacenan y probablemente fabrican las secreciones tóxicas de los Dendrobátidos. La fotomicrografía superior izquierda muestra una sección vertical a través de la parte superior de una de tales glándulas, aumentada 900 veces. La glándula está repleta de grandes gránulos circulares; el conducto de la glándula perfora una capa de células pigmentarias y atraviesa la epidermis y se abre en la superficie de la piel. La micrografía electrónica superior derecha muestra una sección a través de una glándula, aumentada 1500 veces. Las vesículas de los gránulos quizá contengan enzimas implicados en la síntesis de las toxinas. La micrografía inferior muestra gránulos y las vesículas de su interior aumentados 13.000 veces. La micrografía de arriba, a la izquierda, corresponde a una glándula de *Dendrobates tricolor*; las otras, a una glándula de la especie *D. auratus*.

verificaron luego, independientemente, tan curioso uso de las ranas, y se hicieron intentos infructuosos de analizar muestras de veneno raspado de dardos indios. El antropólogo sueco S. Henry Wassén realizó investigaciones sobre el terreno en Colombia en 1934 y en 1955. Wassén determinó que el veneno de las ranas lo utilizaban dos grupos emparentados de indios chocoes a lo largo del río San Juan. Las ranas que obtuvo de los indios, erróneamente identificadas por los zoólogos como *Dendrobates tinctorius*, eran en realidad *Phyllobates aurotaenia* y *P. bicolor*. Consiguió algunos dardos emponzoñados de un grupo aislado de indios chocoes localizado mucho más al sur, en el río Saija. A principios de la década de 1970, nuestro colega Borys Malkin y nosotros mismos recolectamos ranas de esa localidad. Nuestros ejemplares correspondían a una especie no descrita que resultó contener por lo menos 20 veces más veneno que sus parientes del San Juan. Dada la pavorosa toxicidad de la rana le dimos un nombre cuyas connotaciones son evidentes: *Phyllobates terribilis*.

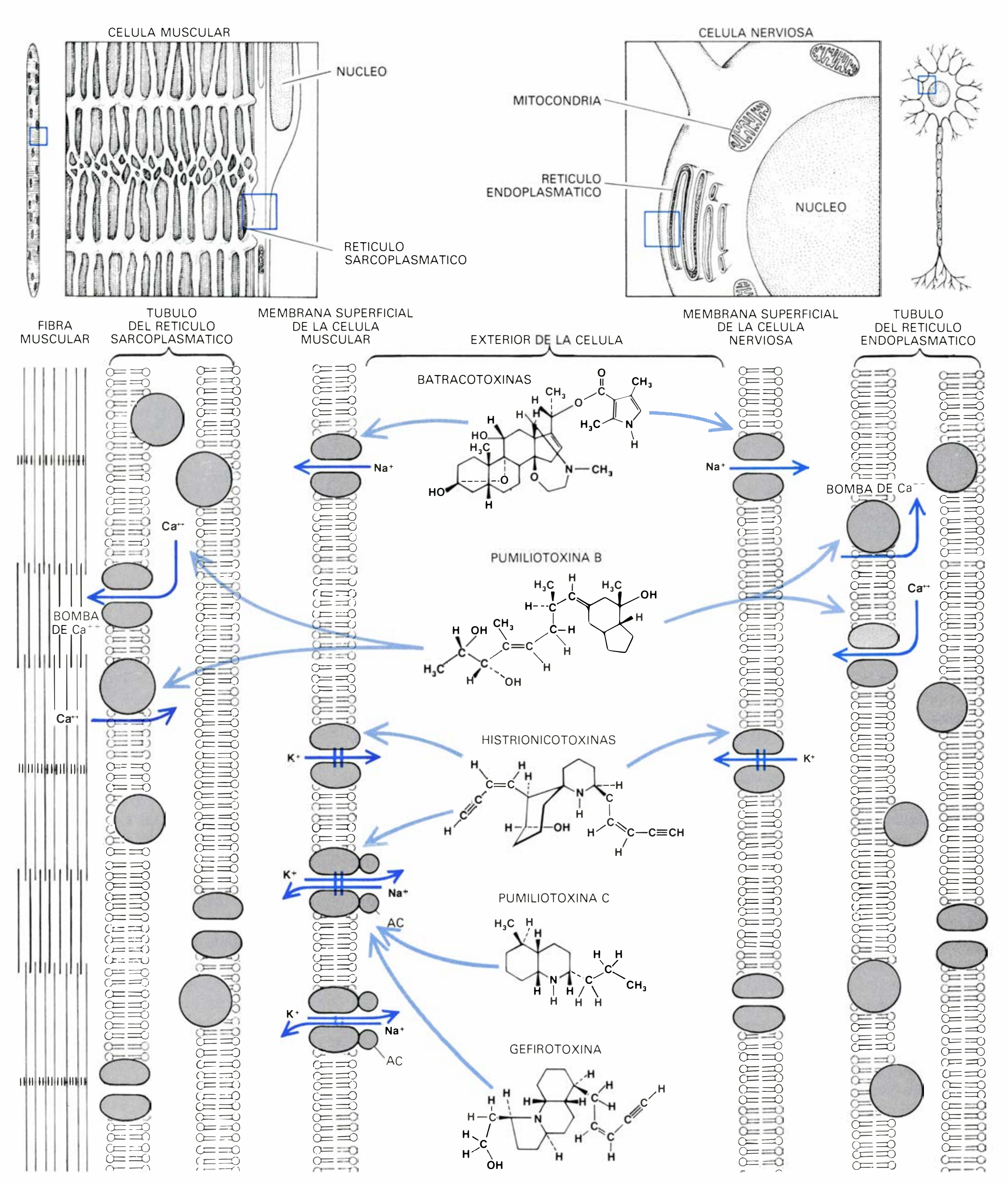
Los indios chocoes del norte emponzoñan sus dardos con las secreciones de *P. bicolor* y *aurotaenia*, de manera muy parecida a como lo describiera Cochrane hace siglo y medio. Se empalan las ranas en un bastoncito y a veces se las acerca al fuego antes de emponzoñar los dardos. Sin embargo, los indios chocoes meridionales envenenan sus dardos frotándolos simplemente sobre el dorso de un ejemplar vivo de *P. terribilis*.

En la actualidad los chocoes emplean cada vez menos la cerbatana; les son más accesibles las armas de fuego. Contrariamente a algunos relatos literarios, las ranas de la familia Dendrobátidos no se aprovechan para envenenar flechas, y su utilización para emponzoñar dardos de cerbatana no parece haber estado muy extendida. Ello es comprensible si se piensa que los tres dendrobátidos más tóxicos se hallan confinados a regiones bastante circunscritas de Colombia occidental. Otras dos especies de *Phyllobates* se encuentran en Centroamérica, pero tienen mucha menos toxina que sus parientes colombianas; además, su área de distribución cae fuera de la zona donde se emplearon dardos emponzoñados, según se sabía.

Los órganos dérmicos que segregan las toxinas de los Dendrobátidos son glándulas "granulares" cuyas minúsculas aberturas se hallan dispersas por la superficie de la piel entre las de las glándulas mucosas de la rana. Las glándulas granulares parecen constituir una característica primitiva común a todas las ranas. Evidentemente, fueron una preadaptación evolutiva conveniente en algunos grupos anfibios de éstos, para la síntesis, almacenamiento y liberación de secreciones biológicamente activas.

Los principios activos de las secreciones dérmicas de los Dendrobátidos son alcaloides, es decir, compuestos nitrogenados cíclicos. Aunque tales compuestos se encuentran con más frecuencia en las plantas que en los animales, los Dendrobátidos constituyen una fuente rica de alcaloides únicos. Hemos recolectado dendrobátidos por toda su área de distribución en América tropical para analizar posteriormente sus secreciones defensivas. Esta labor ha conducido al descubrimiento de una docena de nuevas especies de ranas ponzoñosas y de más de 200 alcaloides nuevos, que representan por lo menos cinco clases distintas de compuestos. La investigación química de los alcaloides de estas ranas se inició en el Instituto Nacional de Artritis, Metabolismo y Enfermedades Digestivas, de los Estados Unidos, por parte de Bernhard Witkop, a principios de la década de 1960. A lo largo de las dos últimas décadas, quienes más han trabajado en esa línea han sido Isabella L. Karle, cristalógrafa de rayos X del Laboratorio de Investigación Naval, y Takashi Tokuyama, químico de la Universidad de Osaka. Los fundamentos de la actividad de cada clase de alcaloides de dendrobátidos revisten un notable interés. Pionero de la investigación farmacológica de estos compuestos ha sido Edson X. Albuquerque, de la Facultad de Medicina de la Universidad de Maryland.

La gran mayoría de alcaloides de dendrobátidos poseen en su estructura molecular un anillo compuesto por un átomo de nitrógeno y cinco átomos de carbono: un anillo de piperidina. En todos los subgrupos de los géneros *Dendrobates* y *Phyllobates* se encuentran compuestos piperidínicos simples. Desde una perspectiva evolutiva, ello prueba que las ranas de estos géneros constituyen una estirpe singular dentro de la familia Dendrobátidos. Sin embargo, en las cinco especies del género *Phyllobates*, la biosíntesis de alcaloides piperidínicos se ha suprimido en gran parte en favor de la síntesis de una nueva clase de alcaloides extraordinariamente tóxicos. Se trata de las batracotoxinas (del griego *batrachos*, rana), que son la



6. SITIOS DE ACCION MOLECULAR de cuatro tipos de compuestos alcaloides procedentes de dendrobátidos; se muestran, esquemáticamente, para el músculo (*abajo a la izquierda*) y para el nervio (*abajo a la derecha*). Las características morfológicas de una célula muscular se muestran arriba a la izquierda y, las de una célula nerviosa, arriba a la derecha. En medio aparecen las fórmulas estructurales de los alcaloides. La batracotoxina impide el cierre de los canales de sodio de la membrana superficial de las células musculares y nerviosas. El flujo de iones sodio despolariza eléctricamente la membrana y detiene la función muscular y nerviosa. La pumiliotoxina *B* parece actuar de dos maneras. En primer lugar, facilita la liberación de iones calcio desde lugares de almacenamiento del interior de la célula muscular, con lo que potencia la contracción muscular. En segundo lugar, inhibe el retorno de los iones calcio a los lugares de almacenamiento, con lo que prolonga la contracción. La pumiliotoxina *B* afecta asimismo a la traslocación del calcio en las neuronas. La histrionicotoxina bloquea a la vez el movimiento hacia el exterior de iones potasio a través de sus canales, situados en la membrana superficial de las células musculares y nerviosas, y el intercambio, en ambos sentidos, de iones sodio y potasio a través de complejos de canales de iones y de receptores de la acetilcolina (AC), situados en la "placa terminal" entre una fibra nerviosa y una célula muscular. El bloqueo de los canales de potasio provoca la contracción muscular y prolonga la liberación de neurotransmisores por parte de las neuronas. El bloqueo de los complejos de canales iónicos/receptores de acetilcolina impide que la AC liberada por los nervios desencadene la contracción muscular. La pumiliotoxina *C* y la gefirotoxina bloquean asimismo el movimiento de los iones a través de los complejos de canales iónicos/receptores de AC, impidiendo que ésta dispare la contracción.

base principal de la ponzoña de los dardos de cerbatana de los chocoes. Estos alcaloides complejos guardan una relación estructural con los esteroides, pero presentan muchas características hasta hoy desconocidas en los compuestos naturales.

Las batracotoxinas figuran entre las más potentes toxinas no proteínicas de la naturaleza. Aumentan selectivamente la permeabilidad de la membrana externa de las células nerviosas y musculares a los iones sodio. Este efecto produce una despolarización eléctrica irreversible de las células, lo que ocasiona en el corazón arritmias, fibrilación e insuficiencia cardíaca. El lugar específico en el que actúa la batracotoxina está asociado a los canales que regulan el flujo de iones sodio a través de la membrana celular. Los canales de sodio desempeñan un papel clave en la conducción de los impulsos eléctricos por las células musculares y las neuronas. La fijación de la batracotoxina impide el cierre normal de los canales. En consecuencia, hay un aflujo masivo de iones sodio y la célula se despolariza. Las neuronas ya no pueden transmitir impulsos y las células musculares permanecen en un estado activado, es decir, contraído. La utilización de la batracotoxina en investigación ha proporcionado importantes conocimientos sobre la función de los canales de sodio.

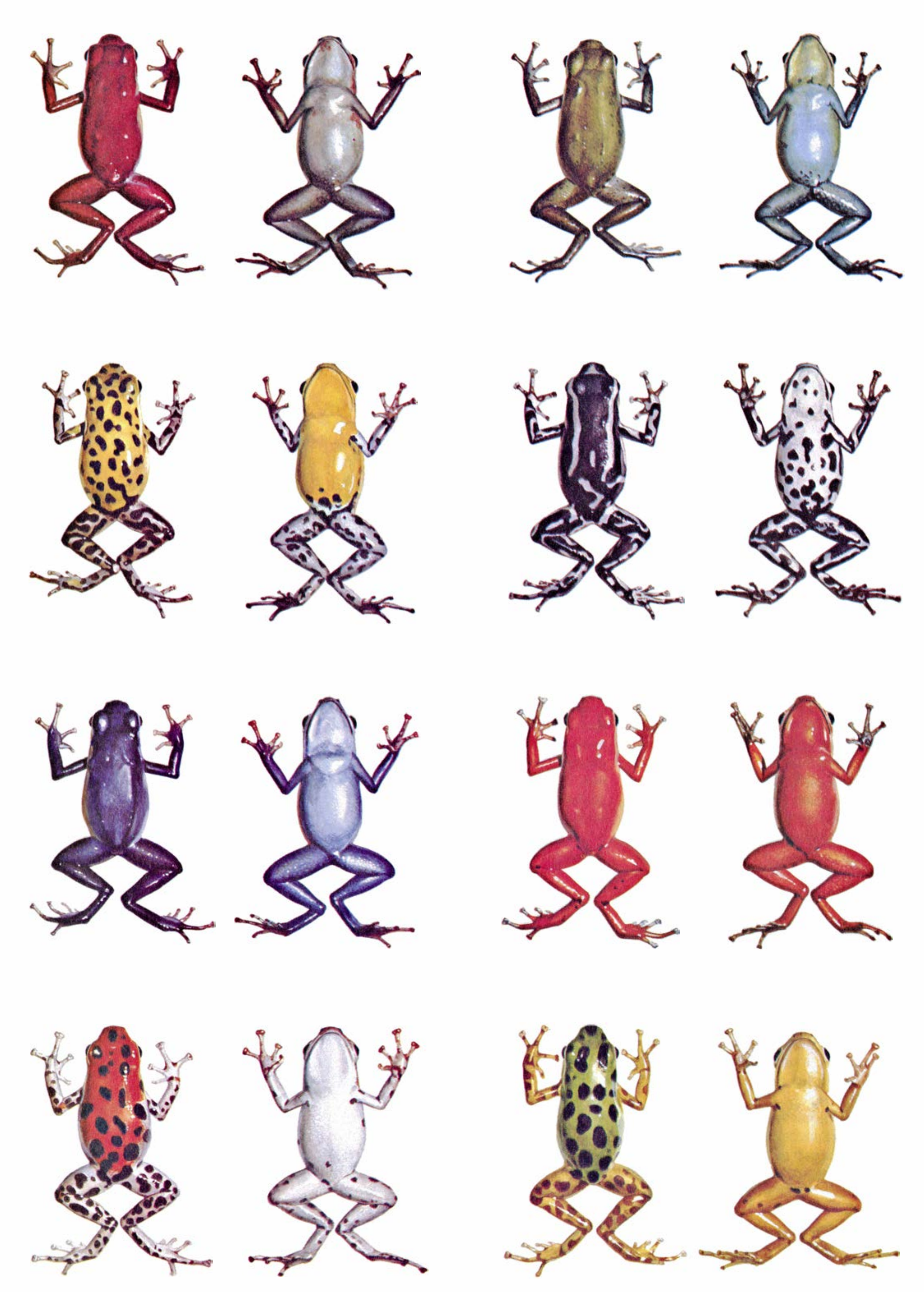
Los otros alcaloides de dendrobátidos son todos de estructura mucho más simple que la batracotoxina y son mucho menos tóxicos, pero no por ello menos interesantes. Muchos de estos alcaloides piperidínicos actúan sobre un canal iónico de la "placa terminal", en la unión entre una fibra nerviosa y una célula muscular. Este canal está asociado con un receptor de la acetilcolina, que es una sustancia neurotransmisora. La interacción del neurotransmisor con su receptor abre en la membrana celular el canal a través del cual pueden fluir los iones sodio y potasio. Si el impulso eléctrico resultante es suficientemente grande desencadenará la apertura de numerosos canales de sodio de la membrana. En las células musculares, este proceso provoca la liberación de iones calcio en el interior de la célula, lo que conduce a la contracción.

Las histrionicotoxinas, alcaloides espiropiperidínicos insólitos que se aislaron por vez primera de la rana *Dendrobates histrionicus*, interactúan con determinados lugares del complejo formado por el canal de la placa terminal y el receptor de acetilcolina, bloqueando de este modo el paso de iones. Ello puede impedir la transmisión de señales desde el nervio hasta el músculo, efecto similar al que ocasiona el curare. Las histrionicotoxinas bloquean asimismo el paso de iones a través de los canales de potasio. Cuando estos canales están abiertos, los iones potasio fluyen hacia fuera de las células musculares y nerviosas, permitiéndoles que retornen a su estado de "reposo" después de la transmisión nerviosa o de la contracción muscular. El bloqueo de tales canales puede alargar la transmisión de mensajes nerviosos y prolongar la contracción muscular.

La pumiliotoxina *B*, un alcaloide indolicidínico único aislado por primera vez de la rana *D. pumilio*, parece afectar al transporte de iones calcio. En las células musculares, el impulso eléctrico hace que se liberen iones calcio a partir de los lugares de almacenamiento internos. Los iones interactúan entonces con proteínas intracelulares del músculo y provocan la contracción. En lugar de bloquear este proceso, la pumiliotoxina *B* parece facilitarlo, y también parece retardar el retorno de los iones calcio a sus lugares de almacenamiento. En consecuencia, la fuerza de la contracción aumenta y se prolonga la contracción, tanto del corazón como del músculo esquelético. Estas acciones tónicas sobre el músculo quizá permitan la aplicación clínica de esta clase de alcaloides.

Las ranas de la familia Dendrobátidos, tengan o no secreciones nocivas, suelen presentar colores vivos. Como ocurre a menudo con los animales nocivos, estas especies son imitadas por algunas otras especies de ranas (mimetismo). Ejemplo de ello es *Eleutherodactylus gaigeae*, que se encuentra junto a *Phyllobates lugubris* y *P. aurotaenia* y en el territorio intermedio entre el de estas dos especies, donde probablemente vivió antaño una forma ancestral de *Phyllobates*. A diferencia de los dendrobátidos a los que mimetiza, la especie no tóxica *E. gaigeae* es activa durante la noche. No obstante, durante el día se la encuentra a veces en la hojarasca donde las especies de *Phyllobates* buscan refugio en su huida. La diferencia en los ciclos de actividad puede, por tanto, no ser de importancia en lo que a los depredadores diurnos se refiere.

El género *Phyllobates* presenta menos variabilidad que el género *Dendrobates*, cuyas especies exhiben un notable polimorfismo de colores y diseños. Muchas especies de *Dendrobates* tienen diseños listados simples, pero otras han desarrollado coloracio-



nes que son uniformemente brillantes o bien manchadas, bandeadas o moteadas de distinta forma. Esta variabilidad puede darse incluso en poblaciones de la misma especie. Sería muy interesante saber de qué modo se origina este polimorfismo y si tiene que ver con el origen de nuevas especies. Una especie en particular, *D. pumilio*, centroamericana, parece presentar un buen argumento con el que responder a estas cuestiones básicas.

D. pumilio es una rana pequeña, que generalmente se encuentra a elevaciones inferiores a los 500 metros en la selva lluviosa del lado caribe de Nicaragua, Costa Rica y Panamá occidental. La variabilidad suele ser pobre en Nicaragua y Costa Rica, donde las ranas son por lo general rojas o de color anaranjado rojizo, con las patas posteriores negras o azul vivo. Sin embargo, en Panamá la variabilidad es extremadamente rica; ello está asociado con la historia geológica de Panamá occidental. Hace unos 12.000 años, el nivel del mar que iba subiendo creó un archipiélago de islas costeras y, en el continente adyacente, una mezcla de bosque lluvioso y de bosque pantanoso de llanura. Antes, probablemente, el archipiélago se había formado y había vuelto a conectarse al continente varias veces. Las poblaciones de *D. pumilio* de la región presentan todos los colores del espectro, desde el rojo al azul; hay incluso una población diseñada en blanco y negro. Las poblaciones parecen especies diferentes; en nuestra opinión, el complejo encierra al menos dos especies, por más que donde más variabilidad se da es en la especie, única según todos los indicios, *D. pumilio*.

El polimorfismo no se debe simplemente a la evolución independiente de poblaciones aisladas en islas; distintas partes de una sola isla pueden estar habitadas por poblaciones muy diferentes. Existen ejemplos de intensa variabilidad dentro de una misma población. Un ejemplo que hemos estudiado a lo largo de varios años se encuentra en el extremo noroccidental de la isla Bastimentos. Allí, el color del dorso de las ranas va desde rojo anaranjado hasta verde metálico, pasando por bronce. En un bosque que cerraba una playa, las ranas de color rojo anaranjado alcanzaban el 96 por ciento de las muestras poblacionales que tomamos cada año, y sólo el 4 por ciento de las muestras presentaba otros colores. En una plantación de cacao situada a la sombra de un bosque y alejada de la playa, sin embargo, las ranas de color rojo anaranjado constituían el 84 por ciento o menos de las muestras, mientras que las ranas de color bronce o verde llegaban al 32 por ciento de los totales anuales. Estas diferencias entre poblaciones de una especie en hábitats distintos son un ejemplo de lo que se conoce por "polimorfismo equilibrado". El fenómeno es, probablemente, la base genética de algunas de las sorprendentes diferencias entre poblaciones de *D. pumilio*.

Aunque las poblaciones de *D. pumilio* que divergen grandemente se hallan por lo general separadas por extensiones de bosque no habitadas por las ranas, encontramos regiones en Panamá continental en las que poblaciones casi puras de ranas rojas se hallaban en contacto con poblaciones igualmente puras de ranas verdes. Las ranas de otros colores escaseaban en ambas poblaciones. Tampoco abundaban en las dos poblaciones otras variantes de color que se parecían a ranas de poblaciones más alejadas. Estos hallazgos indican asimismo que de un acervo génico común pueden surgir poblaciones de ranas de distintos colores.

El problema planteado por *D. pumilio* no se limita a las diferencias de color. Las ranas de las distintas poblaciones varían asimismo en tamaño y en comportamiento. Su hábitat usual puede hallarse sobre el suelo o en los árboles; su respuesta a los depredadores potenciales va desde la conducta tímida y cautelosa hasta la audaz y aparentemente intrépida. Ninguna de estas características parece estar correlacionada con las secreciones defensivas dérmicas de las ranas, cuyas variaciones son asimismo notables. En conjunto, las poblaciones de *D. pumilio* de Panamá occidental producen más de 80 alcaloides, que representan las principales clases de toxinas piperidínicas. El número de alcaloides segregados por ranas de una determinada población va de seis a 24, y muchos de estos compuestos no se han detectado en otros dendrobátidos. *D. pumilio* constituye, entre todos los animales vertebrados, un conspicuo candidato al título de "especie más variable". Pero estamos todavía lejos de comprender los factores que han determinado que así sea.

7. OCHO RANAS, que representan distintas poblaciones de *Dendrobates pumilio*, se muestran en posición dorsal y ventral en la pintura de la página opuesta. Estos ejemplos ofrecen sólo una muestra del extraordinario polimorfismo entre ranas del complejo de poblaciones de *D. pumilio* de la vertiente caribe de Panamá occidental. Todas las ranas se han ilustrado al mismo tamaño, pero en la naturaleza varían tanto de tamaño como de color; difieren, también, en comportamiento y en los alcaloides de sus secreciones cutáneas defensivas.



Bóvedas de abanico

Durante el siglo XIV, los maestros canteros ingleses convirtieron la bóveda nervada gótica en un esqueleto arquitectónico de trazado ondulante. Sus fines fueron estéticos, pero las bóvedas se levantaron con gran solidez

Walter C. Leedy, Jr.

Uno de los problemas fundamentales en arquitectura es cómo encerrar y articular el espacio interno. La situación se complica conforme el espacio va ganando en amplitud, siendo especialmente crítico en una catedral o abadía medieval. En ellas, por la fuerza ascensional de los muros, se destaca claramente la cubierta, la cual, consecuentemente, tenía que dotarse de un diseño llamativo. Además, la cubierta se hacía normalmente de piedra y era muy pesada. La solución adoptada en las iglesias góticas más representativas fue la configuración de la cubierta como una serie de bóvedas, en las que la plementaría de sillares cabalgaba sobre agudos arcos apuntados. Los arcos servían como nervaduras que soportaban, al menos, parte de la carga.

Cuando los constructores medievales lograron mayor experiencia sobre las bóvedas nervadas, descubrieron que la construcción podía facilitarse y el diseño de las mismas hacerse más adaptable, incluyendo nuevos nervios en la estructura de la cubierta. El abovedamiento llegaría a ser un complicado encaje de nervaduras, al que se podía dar multitud de formas. En la Inglaterra del siglo XIV, la proliferación de los nervios condujo a la invención de una estructura completamente nueva: la bóveda de abanico. En ésta, se sustituyó la plementería de la propia bóveda nervada por una estructura redondeada llamada conoide, que se asemeja a un cono dividido en dos a lo largo de su eje, y con su vértice invertido. Cuatro conoides configuraban la unidad modular de la bóveda. Sobre la superficie curva de cada conoide se extendían numerosos nervios, hacia arriba y hacia afuera, hasta alcanzar la cúspide de la bóveda. Precisamente de este esquema en abanico de los nervios tomó el estilo su nombre. Entre los nervios se labraba una complicada tracería.

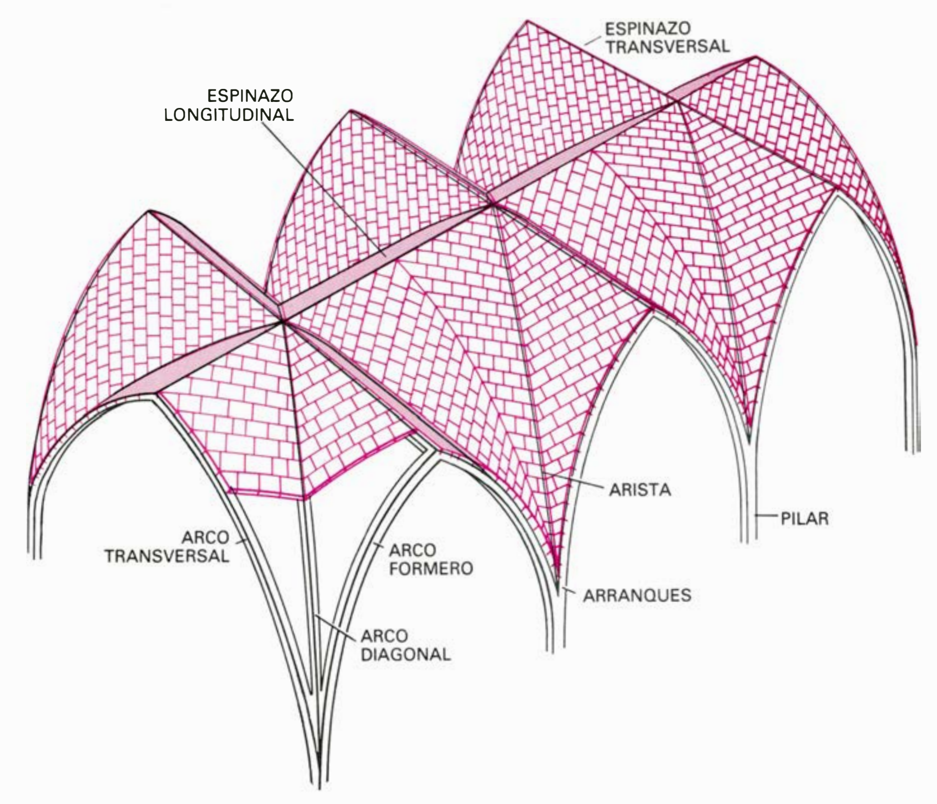
En una bóveda nervada, los nervios están destinados a soportar el peso, pero el conoide de una bóveda de abanico puede actuar como una estructura laminar, en la que las cargas están perfecta y uniformemente distribuidas a lo largo de toda la fábrica de piedra. El constructor medieval no sabía calcular las cargas ejercidas sobre una superficie. A pesar de lo cual, los arquitectos contemporáneos han demostrado que el conoide de una bóveda de abanico es una estructura muy estable. Desde 1350 hasta 1540, aproximadamente, los constructores ingleses sacaron partido a esta circunstancia para levantar más de cien bóvedas de abanico, cubriendo cada vez mayores espacios. Entre las más famosas se cuentan la de la capilla del King's College, de la Universidad de Cambridge, y la de la capilla de Enrique VII, en la Abadía de Westminster.

La elaborada cubierta de un edificio con bóvedas de abanico ofrece un vigoroso encanto visual. La redondez de los conoides se sucede, de uno a otro, a lo largo del plano del muro, produciendo una impresión de volúmenes espaciales airosa y continuamente trazados, en contraste con la angulosidad de las bóvedas nervadas. La intrincada labra de la superficie del conoide destaca su volumen y contribuye a crear una impresión de unidad. Probablemente fuera ese efecto visual el primer objetivo de los constructores que levantaron las bóvedas de abanico. No se puede olvidar, sin embargo, que bajo esa experiencia estética subyacía la solución a un problema arquitectónico fundamental, por el que los creadores superaron con su experiencia práctica la carencia de una mayor cualificación teórica.

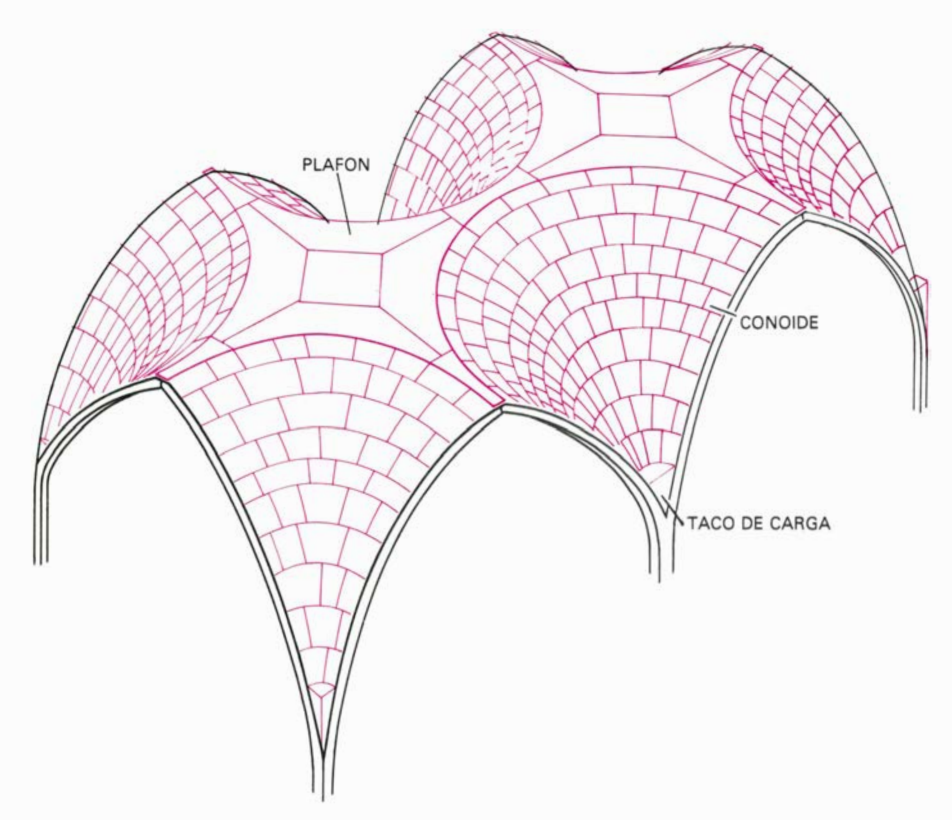
El desarrollo de la bóveda de abanico está íntimamente conectado a las cualidades físicas de la piedra y a las preferencias estéticas de los maestros canteros ingleses y de sus patronos. La piedra puede resistir grandes presiones, pero tiene poca resistencia a la tracción, no aguanta con facilidad que se la tense. La fuerza de la gravedad puede generar cargas importantes en un dintel que sólo se apoya en sus extremos. En una bóveda de cantería, las cargas tienen que convertirse, de algún modo, en fuerzas de compresión. En los arcos apuntados de una bóveda de nervadura, la conversión se logra elevando las dovelas sobre el plano horizontal definido por la línea de impostas. La fuerza de la gravedad se transforma así en un sistema de fuerzas hacia afuera y hacia abajo que se transmite a los muros.

La unidad modular de una cubierta abovedada suele ser un compartimento rectangular, denominado tramo, cuyo eje longitudinal es perpendicular al eje mayor de la sala. En una bóveda de nervaduras, el tramo presenta, generalmente, tres pares de arcos. Los arcos transversales cruzan la sala. Los formeros corren a lo largo del muro en los lados menores del rectángulo y, casi siempre, cobijan los ventanales. Los ángulos opuestos de cada tramo están enlazados mediante los arcos diagonales, cruzados en el centro del rectángulo. Altos pilares, situados bajo los arranques –punto donde se reúnen los segmentos inferiores de los tres arcos–, ayudan a soportar el vuelo de la bóveda. El empuje de los arcos se transmite hasta el suelo por los pilares y por los arbotantes, que se encuentran fuera de los muros, próximos a los arranques.

1. CAPILLA DEL KING'S COLLEGE, de la Universidad de Cambridge, que posee uno de los más hermosos ejemplares de bóveda de abanico. La bóveda principal de la capilla, aquí mostrada, se concluyó en 1515. Es la de mayor amplitud conocida: salva una distancia de 12,7 metros. Las superficies redondeadas que se entrecruzan, los conoides, constituyen la principal estructura portadora de cargas en una bóveda de abanico y actúan como membranas donde las tensiones se distribuyen de manera uniforme. El labrado, derivado de las tracerías de los ventanales, sirve para unificar el interior del edificio.



2. BOVEDA DE NERVADURAS, el típico sistema de cubiertas gótico. Fue concebido por los constructores medievales como una armadura en la que los arcos apuntados funcionan como nervios portadores de los pesos. Cada una de las unidades modulares de la bóveda, denominada tramo, incluye seis arcos dispuestos en rectángulo, con sus intersecciones diagonales. La carga de la bóveda se convierte en una serie de empujes hacia fuera y hacia abajo que se transmiten, en parte, a través de los arcos a los pilares y a los muros. Las superficies de la bóveda situadas entre los arcos y ocupadas por hiladas de sillares se llaman plementos. En Inglaterra, los sillares de los plementos se situaban formando un ángulo de unos 45 grados, desde los espinazos longitudinales o transversales hasta la arista de unión de los plementos.



3. BOVEDA DE ABANICO, que se desarrolló, a partir de la bóveda de nervaduras, en 1350. El conoide es una superficie de rotación, originada por el giro de una línea curva alrededor de un eje vertical. Visto desde el centro del tramo, tiene una sección horizontal convexa y otra vertical cóncava. El conoide actúa por lo tanto como una serie de arcos horizontales y verticales que transmiten empujes al muro.

En la construcción de una bóveda de nervios, los arcos se levantan en primer lugar, empleando una compleja estructura de madera denominada cimbras, que tienen la forma en que se han de colocar los sillares. Cuando los seis arcos de cada tramo se han levantado, sirven de armazón a la plementería, es decir, a las superficies pétreas situadas entre los nervios.

Las bóvedas nervadas, como las que existen en las catedrales de Amiens y Rouen, tienen una apariencia marcadamente lineal, que es el fundamento de la estética gótica. El efecto se lograba interpretando las bóvedas nervadas como un armazón en el que los plementos llenaban el espacio que quedaba entre los nervios. Además, los largos arcos transversales acentuaban la división de la cubierta en tramos. En contraposición a ello, aunque la bóveda de abanico esté todavía compuesta de elementos lineales, la unidad del espacio interior queda más acentuada. La mirada del espectador se ve atraída hacia la bóveda por la reiterada y sugestiva forma de los conoides; la diferenciación entre los tramos tiene generalmente poca importancia visual.

Aunque los principios estéticos de los dos tipos de bóvedas son bastante diferentes, la transición histórica entre ellos no se hizo de una manera brusca. Por el contrario, la bóveda de abanico se desarrolló gradualmente a partir de la construcción de las bóvedas nervadas. Estas eran frecuentes en Inglaterra, en Francia y en otros países europeos, mientras que la bóveda de abanico se desarrolló exclusivamente en Inglaterra. Naturalmente, este hecho plantea la pregunta de por qué las nuevas formas surgieron solamente allí.

La respuesta se halla en los principios estéticos y en la práctica constructiva que prevalecieron en Inglaterra durante la Edad Media. La comparación con Francia es muy instructiva. En Francia, las hiladas de sillares de la plementería se colocaban paralelas a los lados del rectángulo que cubría la bóveda; constituían hileras horizontales desde el arranque de la bóveda hasta su punto más alto. Cada cuarto de tramo semejaba un tejado a dos aguas; el espinazo de cada cuadrante se denomina transversal o longitudinal, según su orientación. A lo largo de la arista que une los cuadrantes, la plementería se colocaba paralela a ambos espinazos, formando ángulos rectos con los lados del tramo. Durante cierto tiempo, los constructores ingleses siguieron el procedimiento francés. Sin embargo, muy

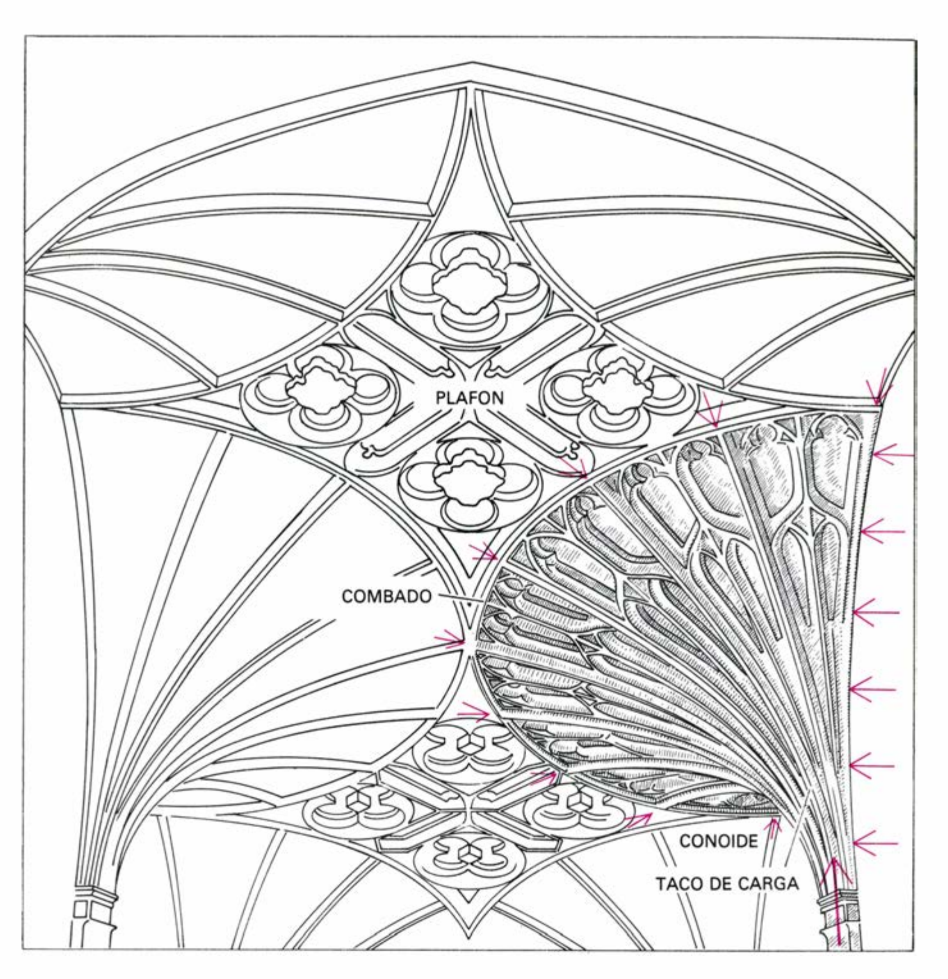
pronto comenzaron a situar las hiladas desde el espinazo del cuadrante hasta la arista formando ángulos de casi 45 grados respecto a ambos espinazos. Esta peculiar forma de situar las hiladas llevó, en buena medida, a que los conoides se concibieran como estructuras redondeadas.

El método inglés también se diferenció en otros procedimientos. En las bóvedas francesas, la plementería se situaba sobre los nervios, que sólo quedaban unidos por la capa de mortero. En Inglaterra, la parte superior del nervio estaba profundamente dentellada; la plementería se encajaba en los dentellones antes de unirla con cemento. Esta práctica, llamada hoy ensamblaje, tendía a hacer de nervio y bóveda una unidad. He aquí la prueba arquitectónica de que el constructor inglés consideró la nervadura como un elemento portador de cargas.

También el diseño inglés de los arcos difiere del utilizado en Francia. Las cualidades arquitectónicas más altamente cotizadas en este último país eran la iluminación, la altura y una estructura diáfana. Para aumentar la superficie de los ventanales, los arcos formeros del muro se peraltaban con frecuencia para elevar la base del arco por encima de los arranques, en el muro baquetonado. El peralte tuvo consecuencias importantes. La plementería próxima a la ventana tuvo que alterarse para acomodarla al perfil del arco. Además, se redujo la superficie del muro entre los ventanales y se concentró en un estrecha faja de muro el empuje hacia afuera de la bóveda. Hubo que situar adecuadamente los arbotantes para evitar el derrumbamiento del muro.

En Inglaterra había menos interés por la altura y la diafanidad. Los muros eran generalmente más bajos y más gruesos. Los ventanales eran más pequeños y, por tanto, los arcos formeros del muro no tenían que peraltarse. Este aumento de la superficie del muro supuso una mejor distribución de los empujes de la bóveda y una menor necesidad de contrafuertes.

La bóveda de nervaduras se desarrolló enormemente en la Europa del siglo XII. Los constructores medievales aprendieron pronto que la adición de otros dos tipos de nervios podía aumentar la flexibilidad del diseño de la bóveda. En el primer tipo, los terceletes corrían de los arranques al espinazo de la bóveda a modo de bisectriz, dividiendo el ángulo constituido por los nervios diagonales y los transversales o los espinazos. Las ligaduras eran cortos miembros estructurales insertados en ángulo entre los nervios mayores.



4. EL CONOIDE DE UNA BOVEDA de abanico debe estar comprimido a lo largo de todos sus márgenes para mantener el equilibrio. El dibujo muestra la bóveda de la iglesia de Santa María de North Leigh, en el condado de Oxford, terminada hacia 1440. El pesado plafón situado entre los conoides ejerce una fuerza de compresión a lo largo del combado y sirve de clave a toda la bóveda. El taco de carga, construido fuera del muro, en el ángulo del tramo, soporta el vértice del conoide. Los muros suministran un empuje hacia dentro, que se opone a los empujes transmitidos por los conoides. Los nervios labrados sobre la superficie cumplen un papel fundamentalmente estético: no son elementos estructurales.

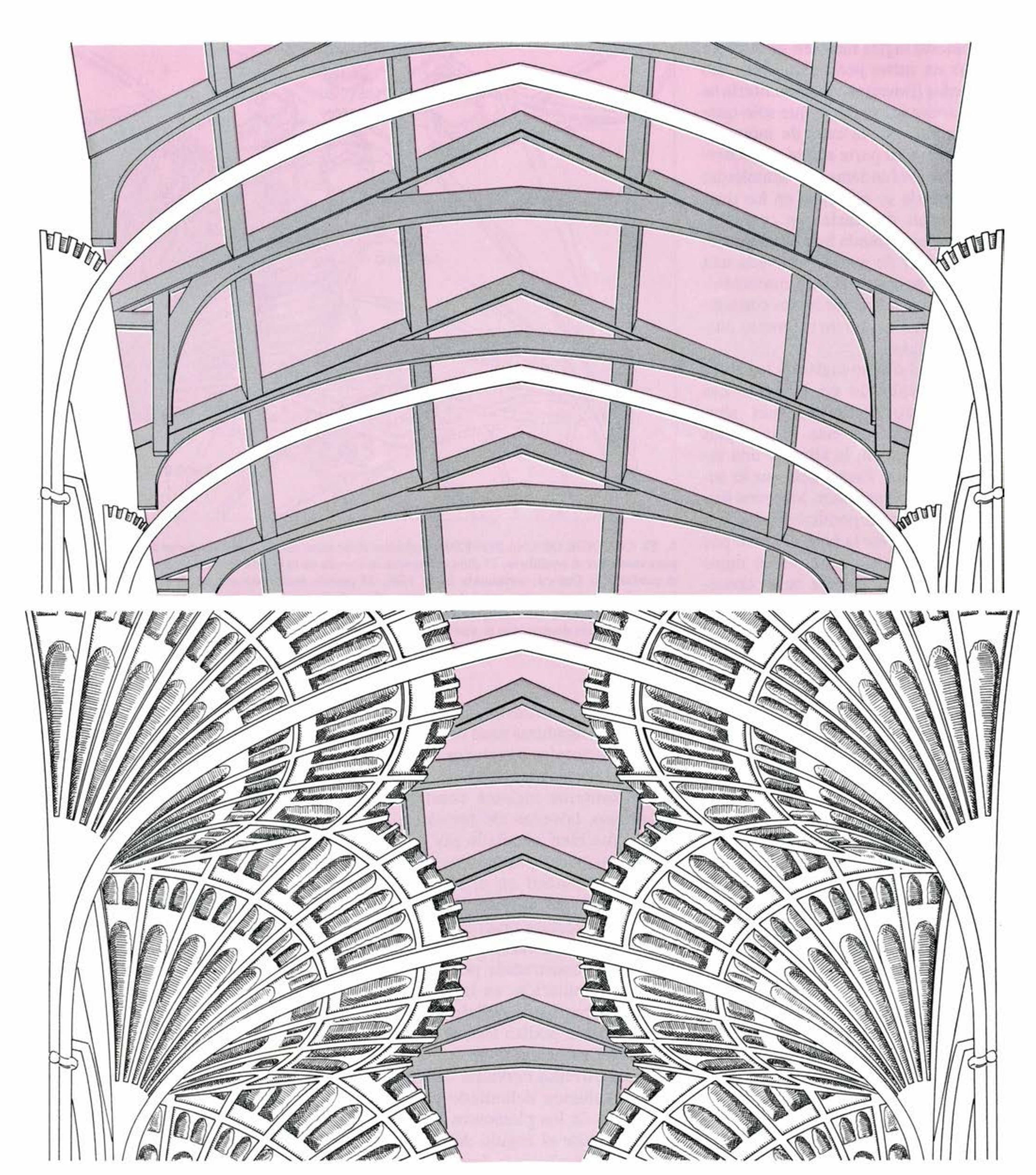
Los canteros ingleses construyeron numerosas bóvedas de terceletes. Su diseño fue bien asimilado por el estilo inglés, en parte porque permitía una mayor flexibilidad en el ajuste de la curvatura de los nervios. En contraste con la arquitectura practicada en Francia, en donde la curvatura de los nervios estaba constreñida por el afán de altura e iluminación, en Inglaterra, los terceletes, partiendo radialmente desde los arranques, podían formar un llamativo esquema geométrico.

En las bóvedas nervadas del alto gótico, el volumen delimitado por la intersección de los plementos a lo largo de la arista en el ángulo del tramo era bastante irregular y la forma se determinaba por otras consideraciones del diseño. Cuando se añadieron más nervios a la bóveda de terceletes el volumen se hizo más regular y empezó a presentar una forma con interés propio. A la vez, la manera inglesa de colocar las hiladas empezó a producir un curioso resultado. Como he señalado, en Inglaterra las hiladas de los plementos hacían ángulo entre los espinazos longitudinales o transversales y la arista. En una bóveda de terceletes, las hiladas de los plementos, en los ángulos de los tramos, empiezan a aproximarse a la forma de polígonos regulares y uniformes. Los polígonos eran concéntricos en los arranques de la bóveda y aumentaban de tamaño hacia la cúspide de la bóveda.

Durante la primera mitad del siglo XIV, las principales condiciones previas para la construcción de una bóveda de abanico se recogían en las bóvedas de terceletes de hiladas concéntricas y nervios rebajados. Faltaban dos avances técnicos para que llegara a plasmarse la bóveda de abanico. El pri-

mero era la práctica de labrar un segmento del nervio y la zona inmediata del plemento en el mismo sillar. Las piedras así labradas se colocaban entonces juntas para formar la bóveda, de la misma manera que se sitúan los bloques de hielo para levantar un iglú. Este tipo de construcción, que se llama de sillería articulada, se empleó en muchas bóvedas de abanico.

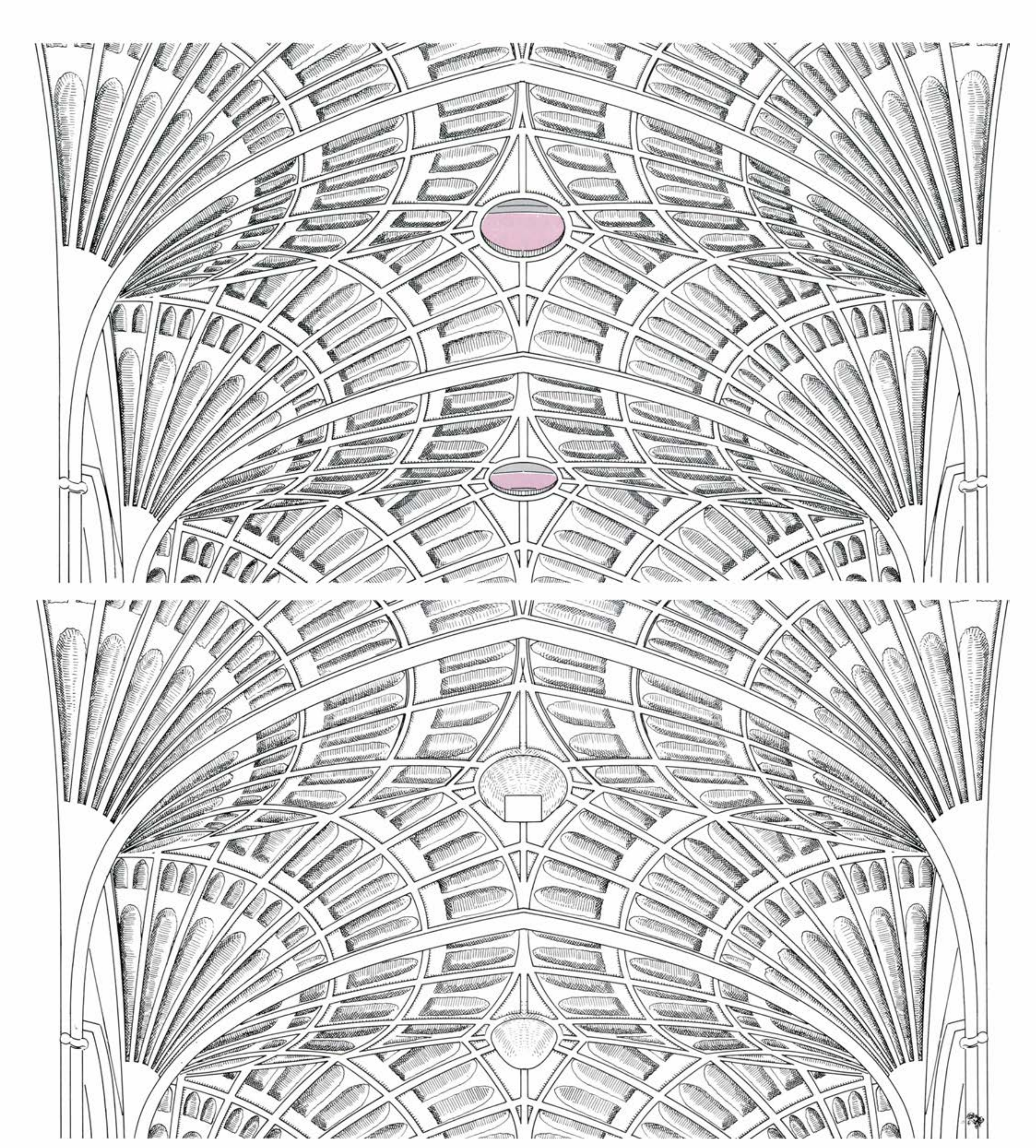
El uso de la sillería articulada por parte de los canteros ingleses no parece obedecer razones fundamentalmente técnicas. Debióse, por contra y en mayor medida, a inquietudes estéticas. Los constructores ingleses valoraban en alto grado la articulación de la superficie de la bóveda. Tal articulación es más fácil de lograr con largos bloques sólidamente acoplados que con las nervaduras y los pequeños plementos in-



5. CONSTRUCCION DE LA BOVEDA de la capilla del King's College, que comprendió cuatro etapas. Los sillares, traídos en barco por el río Cam, o por carretera, se labraron en talleres situados próximos al pie de la obra. Parece que se asentaron con sus superficies decorativas sólo parcialmente terminadas. Se utilizó una serie de cimbras para configurar la bóveda durante la construcción. Tras la conclusión de muros y tejado se levantaron los arcos transversales que dividen la bóveda en tramos (*arriba a la izquierda*). Se pusieron entonces los conoides, procediendo hacia adentro, desde los arcos trans-

tercalados. Cuando el tamaño de los sillares fue aumentando para obtener una superficie articulada, los sillares tendieron, naturalmente, a asumir funciones estructurales.

La introducción de la sillería articulada condujo a otra innovación: en los nervios, el eje mayor de su sección transversal se hizo perpendicular a la superficie de la bóveda. En una bóveda de nervaduras, el eje mayor caía perpendicular al suelo. Cuando la sillería articulada se emplea en una bóveda de abanico, sólo se puede formar, de manera perfecta, un nervio con sillares diferentes si éstos son perpendiculares a la superficie de la bóveda. La proliferación de los nervios y su consiguiente intersección los convirtió en una cuestión de tipo estético cada vez más importante.



versales hacia el centro del tramo (*abajo a la izquierda*). Los conoides se formaban por hiladas sucesivas. Cuando los arcos horizontales de unión de los conoides se encontraban en el centro del tramo, se colocó en su lugar el amplio sillar transversal que servía de clave de ambos arcos. Se situaron entonces los sillares del espinazo longitudinal (*arriba a la derecha*). Finalmente se instaló en su lugar, desde arriba, el amplio florón del centro de cada plafón, que llegan a pesar hasta 1400 kilogramos cada uno (*abajo a la derecha*). Tras retirar las cimbras se procedió a dar la labra final a los nervios y a la tracería.

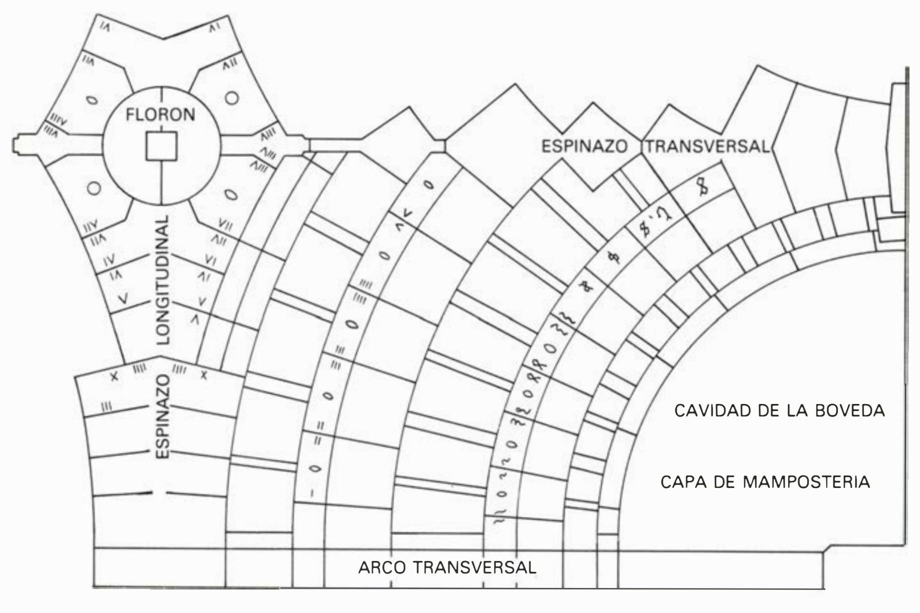
A mediados del siglo XIV se daban ya en Inglaterra la mayor parte de los elementos de la bóveda de abanico. De acuerdo con las investigaciones más recientes, éstas son las características definitorias de una bóveda de abanico: bóvedas conoidales de formas geométricas regulares; nervaduras situadas regularmente, todas con la misma curvatura; un plafón, es decir, un panel central, independiente; nervios perpendiculares a la superficie de la bóveda y superficie labrada geométricamente. Esta serie de cualidades parecen haber surgido, por vez primera, en los pequeños doseles cuadrados de las tumbas de los grandes nobles. (El origen del estilo quizá no pueda nunca determinarse con seguridad, debido a la destrucción de muchos monumentos durante el reinado de Enrique VIII.) Habida cuenta de su pequeñez, se consideran a menudo, las bóvedas de las tumbas, cual miniaturas o variantes decorativas de las bóvedas de abanico. La aparición de éstas bien podría fijarse en la capilla de la Trinidad de la abadía de Tewkesbury, en el condado de Gloucester, concluida en torno a 1380. La mayor bóveda de abanico, y uno de los primeros y más significativos ejemplos de este tipo de abovedamiento, se encuentra en el claustro de la catedral de Gloucester, que se comenzó en la segunda mitad del siglo XIV.

Un rasgo esencial de las bóvedas de abanico es que los conoides son superficies de rotación, es decir, que su forma surge por la rotación de una línea curva alrededor de un eje vertical. El eje es el ángulo del tramo y, la línea curva, la nervadura. A consecuencia de ello, la sección horizontal del conoide es un círculo, o una porción del mismo, en cada nivel. El círculo es convexo con respecto al centro del tramo. Los nervios, por el contrario, son cóncavos con respecto al mismo. De ahí que, en cada punto de la superficie de un conoide, una línea horizontal convexa intercepte otra vertical cóncava.

Por ser circular la prolongación del conoide desde el ángulo del tramo, la bóveda de abanico se acomoda mejor a un tramo cuadrado. (También se construyeron bóvedas con formato rectangular, pero se requería gran habilidad para adaptar la forma al tramo.) En un tramo cuadrado, el conoide tiene una sección horizontal que es un cuarto de círculo. En el centro de la bóveda queda un espacio, aproximadamente romboidal, constituido por la intersección de los cuatro conoides. Este espacio lo ocupa un plafón. En el límite curvilíneo entre el plafón y el conoide, los constructores insertaron a menudo un combado para dividir visualmente las dos estructuras.

El constructor medieval no era formalista. Se sentía libre para alterar los elementos de una bóveda de abanico y así acomodar los requerimientos del diseño a su propio gusto. Por ejemplo, en la capilla del King's College, los nervios verticales parecen estar a intervalos iguales; en realidad, hay pequeñas discordancias, derivadas del hecho de haberse decidido la construcción de la bóveda de abanico cuando el edificio había sido parcialmente levantado. La abadía de Sherborne, en el condado de Dorset, presenta una hermosa bóveda de abanico que ejerció notable influencia entre los creadores del siglo XV. Los conoides del presbiterio presentan secciones horizontales que son más poligonales que circulares. Además, los conoides están construidos por nervios separados y pequeños plementos, y no siguiendo el sistema de sillería articulada. A pesar de todo, ambas bóvedas, las del King's College y las de la abadía de Sherborne, pueden describirse como bóvedas de abanico, debido a su apariencia, no obstante sus divergencias con respecto a la configuración de la bóveda de abanico ideal.





6. MARCAS DE SITUACION de los sillares de la bóveda de la capilla del King's College. Permitían a los asentadores efectuar las labores del abovedamiento guiándose sólo por las marcas. Los símbolos se labraban en los talleres, donde se prefabricaba la bóveda. La marca circular del centro del sillar indica el cuadrante del tramo en el que se tenía que situar. Las marcas de los extremos, próximas a las uniones, son los numerales arábigos 3 y 4, que indican la hilada del conoide en que debía situarse el sillar, su posición en la hilada y el momento de su inserción. Las marcas de situación se aprecian aún sobre la superficie superior de los sillares de la bóveda. En el esquema, un cuadrante de un tramo de la bóveda principal.

La construcción de las bóvedas de abanico puede clasificarse en tres períodos, comenzando con la construcción de las bóvedas en miniatura de los doseles de las tumbas. El primer período finaliza con la conclusión del claustro de la catedral de Gloucester, en 1412. Entre 1412 y 1430 no se construyeron bóvedas de abanico de mayores proporciones. Esta fue una época

de penuria, con altos impuestos y depresión económica; las bóvedas de abanico eran costosas y es probable que ni los particulares ni las instituciones pudiesen contar con el capital necesario para emprender la construcción. El diseño de la bóveda del presbiterio de la abadía de Sherborne, a fines de 1430, marca el comienzo del segundo período, que finaliza hacia 1475. En el tercer período, desde 1475 hasta casi 1540, se construyeron muchas de las mayores y más importantes bóvedas de abanico.

Una abrumadora mayoría pertenecen a edificios religiosos y un considerable número se levantaron en capillas particulares. Una capellanía era una fundación establecida para pagar las misas destinadas al alma del fundador; en la capilla particular se celebraban estas misas. La mayor parte de estas capillas particulares eran dotadas por los nobles, quienes intentaban por igual llamar la atención sobre la grandeza del fundador y asegurar la salvación de su alma. Las bóvedas de abanico, por su evidente impacto visual y su intrincada tracería, se acomodaban perfectamente a este propósito.

Durante la segunda mitad del siglo XV, el gusto por las bóvedas de abanico alcanzó por igual a la clase media y al rey. La capilla de San Jorge del Castillo de Windsor, que es una fundación real, tiene bóvedas de abanico que datan de 1480. Poco después, en 1500, comenzaron los trabajos de la más magnífica entre todas: la correspondiente a la capilla de Enrique VII, en la abadía de Westminster, que el rey Enrique había concebido como una inmensa capilla de los Tudor. Por consiguiente, la Iglesia, la nobleza, la clase media y la corona, todos contribuyeron a la difusión de esta nueva forma arquitectónica.

El desarrollo de la bóveda de abanico fue acompañado por un aumento en la escala de los proyectos. La bóveda de la capilla de la Trinidad de la abadía de Tewkesbury, finalizada hacia 1380, salva una distancia de 1,7 metros. La bóveda principal de la capilla del King's College, que se terminó en 1515, franquea los 12,7 metros. Esta es la mayor extensión de una bóveda de abanico. Una de las razones por las que los constructores ingleses tardaron más de siglo y medio en alcanzar la máxima extensión de una bóveda es que no tenían mucha idea de cómo trabajan las de abanico.

Los arquitectos del siglo XX, utilizando avanzados procedimientos matemáticos, han demostrado que, para que el conoide de una bóveda de abanico esté en equilibrio, es necesario que se apoye a lo largo de todos sus bordes. Un peso importante debe aportar una fuerza de compresión en la parte central del combado que separa el conoide del plafón. La carga es aportada por el propio plafón, que es un pesado sillar en forma de plato. Los grandes florones o sillares prominentes y decorados que constituyen el centro de los plafones en la capilla del King's College pesan unos 1400 kilogramos cada uno.

El plafón sirve como clave de los arcos verticales del conoide. Al igual que en los arcos nervados, los empujes se dirigen hacia fuera y hacia abajo. Los empujes hacia abajo se transmiten al fondo del conoide, donde la estructura carga contra el muro. En este punto, el conoide es soportado por el taco de carga, un saliente construido fuera del muro en el ángulo del tramo. Como se señaló antes, el conoide está compuesto por hiladas concéntricas ascendentes a partir de los arranques; cada hilada adquiere la forma de un arco cuyo trazado es aproximadamente horizontal. Los empujes hacia afuera se transmiten a través de los arcos horizontales al muro, el cual origina un empuje en sentido contrario. El conoide está, pues, comprimido entre el plafón, los muros y el taco de carga.

Las tensiones tienden a distribuirse, en el conoide, bastante equitativamente, más que a concentrarse en las nervaduras. La forma específica en que se distribuyen las tensiones tiene importantes implicaciones para la estabilidad del conoide que, como ya he señalado, es una membrana arquitectónica. Una membrana arquitectónica se define como una estructura mucho más delgada que ancha, sometida a tensiones. La acción de los empujes sobre cada punto de la membrana puede representarse por un vector con componentes hacia fuera y hacia abajo; el conjunto de todos los vectores forma una superficie imaginaria denominada superficie de empuje. Análisis matemáticos demuestran que, para que el conoide esté en equilibrio, la superficie de empuje debe descansar dentro de la superficie física de la membrana del conoide. Si la primera sobresale de esta última es probable que la estructura se derrumbe.

Los constructores de bóvedas de abanico no tenían medios para conocer este hecho fundamental. El análisis de los empujes en una membrana trascendía su capacidad teórica. No obstante, fundados en su larga experiencia, los constructores desarrollaron varias técnicas que aumentaron considerablemente la estabilidad de la bóveda. El estrecho volumen que queda en la base del conoide, sobre el taco de carga, es donde finaliza la cavidad de la bóveda. En muchas bóvedas de abanico esta cavidad se rellenó con mampostería y cemento hasta alcanzar un altura cercana al metro. En la capilla del King's College la mampostería alcanza exactamente la misma altura en cada conoide, demostrando que fue un elemento previsto en la estructura y no un hecho casual.

La presencia del relleno de mampuesto tiene tres consecuencias importantes. Primero, que los empujes de la superficie se extienden al relleno de mampostería, el cual transmite dichos empujes al muro. De este modo, los empujes se distribuyen sobre una superficie mayor de muros, cosa que no ocurriría si esta operación la realizase solamente la membrana. Segundo, la mampostería actúa como una carga volante sobre el suelo. Esta carga es contraria a los empujes hacia fuera del conoide, alcanzando los muros en el ángulo del tramo; allí reduce, además, la tensión ejercida sobre los muros.

Más significativo para la estabilidad del conoide es el hecho de que el relleno de hormigón acorta el vuelo de la bóveda, que funciona como membrana arquitectónica. El mampuesto no puede actuar de membrana, porque es un cuerpo en tres dimensiones y no una lámina delgada. Puesto que los empujes de la bóveda pasan a la capa de mampostería, la parte inferior del conoide no actúa como membrana. Jacques Heyman, de la Universidad de Cambridge, ha demostrado que el sector de la bóveda que funciona como tal es la parte situada entre el límite de la capa de mampuesto y la parte inferior del plafón. Esta distancia puede ser bastante corta con relación al espacio total que cubre la bóveda. Disminuyendo la extensión de la membrana, aumenta la proporción entre grosor y largura. A una membrana más corta y gruesa corresponde una mayor probabilidad de que la superficie de empuje descanse entre las superficies del conoide.

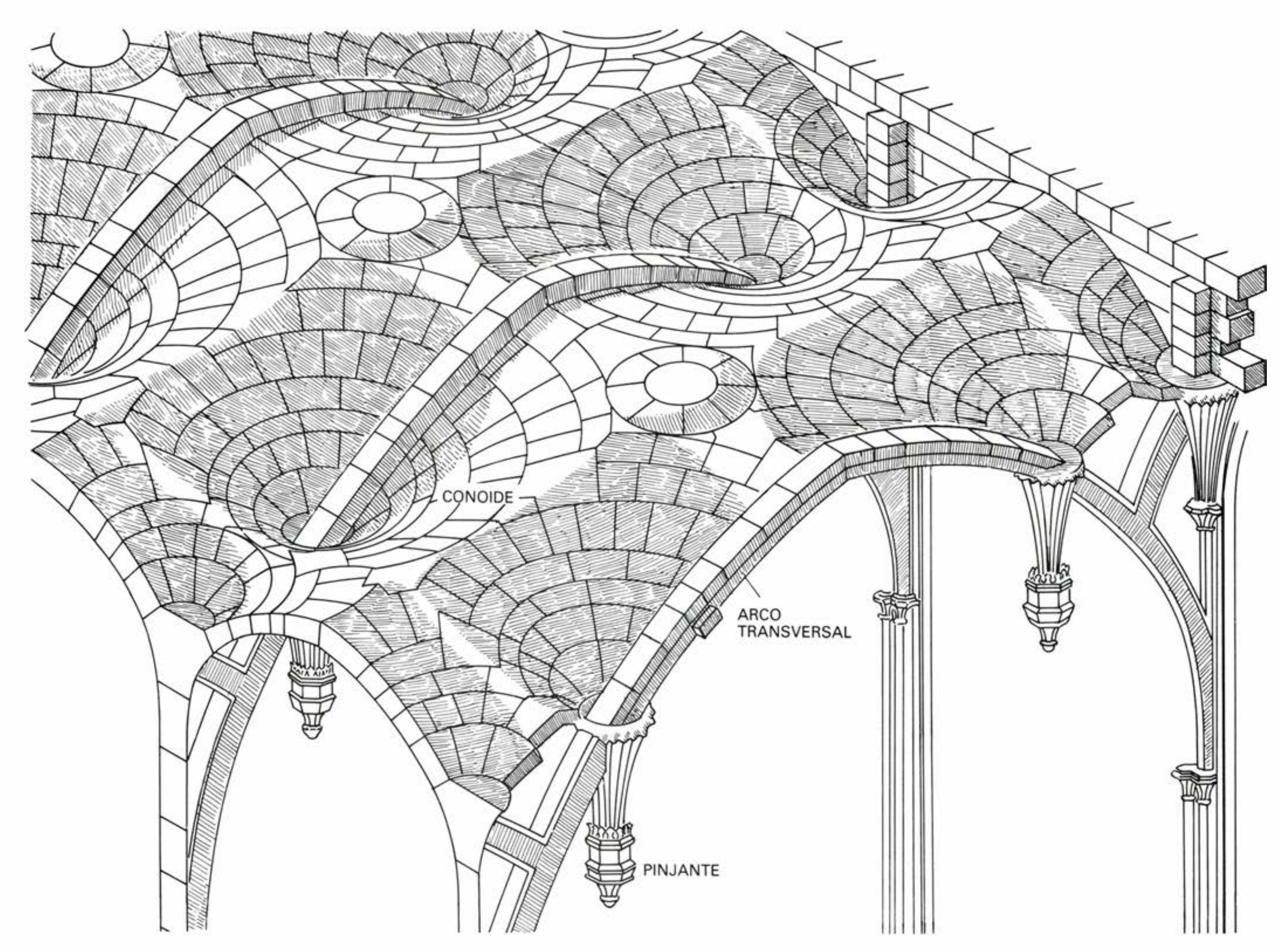
De este modo, sin sobrados conocimientos teóricos, los constructores se las arreglaron para dotar a la bóveda de abanico de estabilidad. Si ésta quedaba asegurada, el otro riesgo a considerar sería el del derrumbe por compresión, es decir, por la fractura

del material de construcción. Mas, por tener los sillares una tremenda fuerza de compresión, no era probable que ocurriera un fallo de este tipo, debido al tamaño de las bóvedas. Por eso, las bóvedas de abanico de mayores proporciones podían construirse siguiendo los mismos principios empleados en las de menor tamaño. Sin embargo, los constructores tardaron muchas décadas en entender que la estabilidad de una bóveda de abanico no dependía de sus proporciones. A finales del siglo XV se reconoció esta circunstancia y se comenzaron a ensayar abovedamientos bastante audaces.

La construcción de bóvedas de abanico de grandes proporciones precisó nuevas técnicas. La piedra para la bóveda de la capilla del King's College fue llevada en barco por el río Cam o en carretera hasta el pie de la obra. Las piedras desbastadas en las canteras formaban largos bloques; un sillar típico sería como un cubo de unos 75 centímetros de lado. Una serie de talleres provisionales se levantaron al pie de la obra y la sillería fue parcialmente labrada en ellos.

En realidad, en los talleres se prefabricó la bóveda de la capilla. Los sillares desbastados se labraron con la forma y proporciones adecuadas, dándose un perfecto acabado a las superficies de unión. Por el contrario, las nervaduras y la tracería que figurarían sobre la superficie del conoide sólo precisaron un labrado somero. Los sillares se marcaban con un símbolo, que señalaba el cuadrante del tramo donde debían situarse. En el borde del sillar, cerca de la juntura, se grababan números arábigos o romanos para indicar el nivel horizontal del conoide asignado, su lugar dentro del mismo y el momento exacto en que debía incorporarse al proceso constructivo.

El asentador tenía sólo que observar las marcas para colocar los sillares en el lugar correcto. La prefabricación de los sillares de la bóveda puede haber requerido más habilidad que la construcción en sí misma. De un estudio de las marcas de construcción que son visibles en las superficies de los conoides de la capilla del King's College deduzco que los tramos de la bóveda principal se construyeron así. Primero se levantaban los muros y se instalaba la techumbre, siguiendo la práctica habitual en los edificios medievales. Después se construían los largos arcos transversales situados entre los tramos. Para la colocación de las dovelas probablemente se precisaron complicadas cimbras. El paso siguiente fue la construcción de los conoides, comenzando por los lados mayores de cada tramo, a lo largo de los arcos transversales, y siguiendo hacia dentro. En cada conoide se colocaba, al mismo tiempo, una hilada. Cuando los arcos se encontraban en el centro del tramo, se colocaba el voluminoso sillar, que era clave de los arcos



7. LA CAPILLA DE ENRIQUE VII, de la abadía de Westminster, cuenta con la más extraordinaria de las bóvedas de abanico. Los trabajos de abovedamiento comenzaron en torno a 1500. El ingenioso diseño de la bóveda combina la estructura en armadura y la estructura de membrana. Robustos arcos transversales soportan los pinjantes situados cerca del muro del tramo. Los conoides están construidos sobre los pinjantes. En ellos, los arcos transversales atraviesan la bóveda y continúan por encima de la misma, por lo que, en la parte central de la bóveda, dejan de verse los arcos. La mayor parte de la carga la recibe el arco al pasar a través del conoide. El arco funciona como un miembro de la armadura y el conoide como una estructura de membrana.

horizontales. Una vez finalizados los conoides, se ponían en su lugar los sillares longitudinales de los espinazos, a cada lado del florón central. El paso final era la inserción de este último.

Los florones se colgaban desde arriba. Dado su gran peso provocaban una compresión de la bóveda, que podía cambiar ligeramente su forma y a la vez permitía derribar rápidamente las cimbras. Cuando estas se habían desmontado, se realizaba, probablemente, la labra final de la superficie del conoide. Después, como podían haber aparecido algunas grietas, se recogían los conoides con mortero y, en ocasiones, se pintaban. La bóveda de la capilla del King's College estaba preparada para ser pintada y dorada.

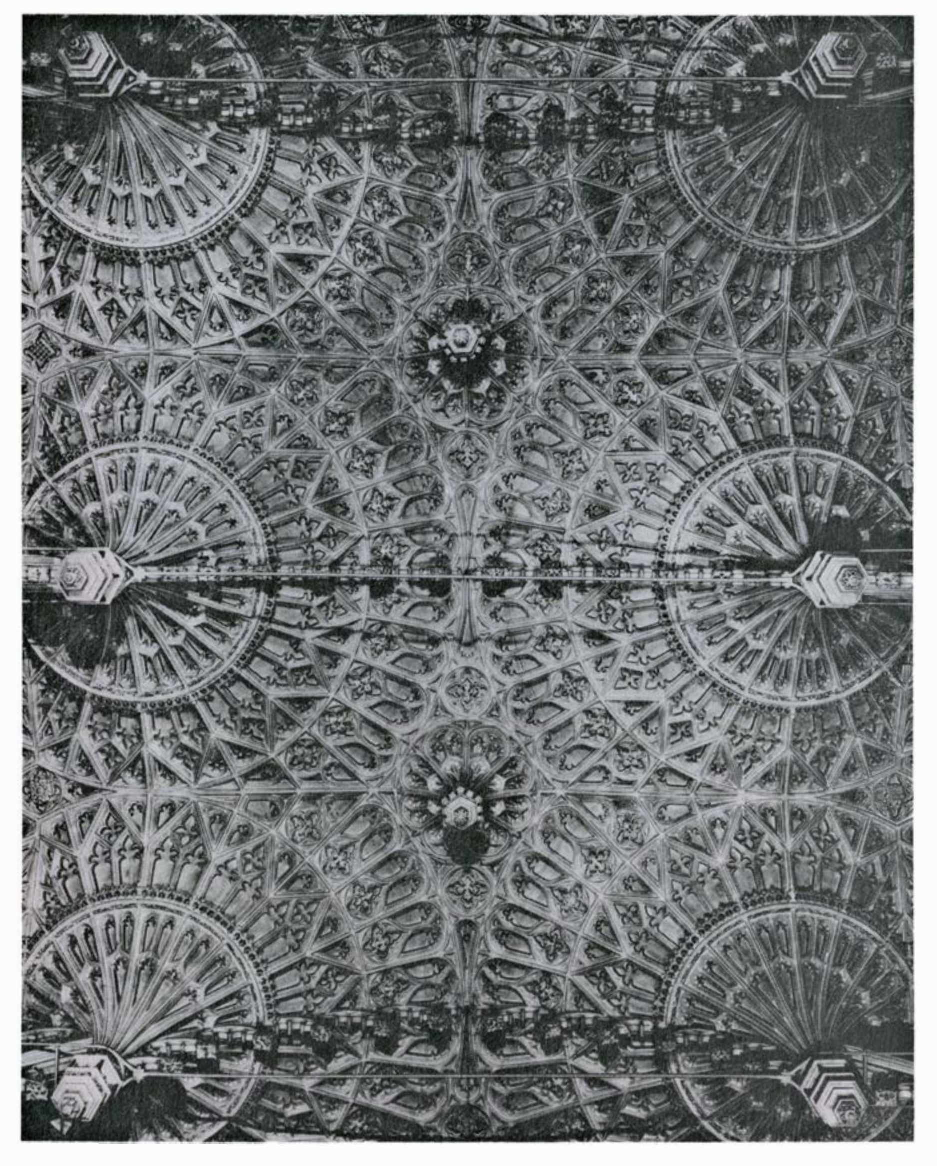
Aunque la bóveda principal del King's College es la mayor de todas las de abanico existentes, la más impresionante quizá sea la de la capilla de Enrique VII. Mide 10,6 metros de amplitud, algo menor que la del King's College y posee un diseño asombroso, en el que se combina una simplicidad de formas tridimensionales con una estructura superficial compleja y regular. Debido al alto grado de superficie de articulación, que se consideró adecuada para un monarca, la bóveda se construyó enteramente en sillería articulada.

La solución estructural de la capilla de Enrique VII es ingeniosa. Amplios arcos transversales, ocultos en parte de su desarrollo, constituyen el estribo de los pinjantes situados cerca del muro. Los conoides se construyeron abriéndose por encima de los pinjantes. En los pinjantes, el arco transversal atraviesa el conoide de la bóveda y continúa por encima de la misma, desapareciendo a la vista en la parte central del tramo. Prácticamente toda la carga estructural se transmite a los arcos transversales en los pinjantes. El arco conduce los empujes hasta los muros y los contrafuertes. De este modo, en la bóveda se combinan la estructura de membrana (los conoides) y la estructura de bastidor (los arcos transversales).

A diferencia de los conoides de muchas bóvedas de abanico, en las que cada sección horizontal es sólo un segmento de círculo, el conoide central de la capilla de Enrique VII tiene por sección horizontal un círculo completo. Este hecho es simbólicamente adecuado porque el círculo, representando la bóveda celeste y el disco solar, era un símbolo importante en la política iconográfica de los Tudor. En 1500, cuando se diseñó la capilla de Enrique VII, los Tudor eran una joven y engreída dinastía. Enrique había tomado el poder hacía sólo quince años, sobre la base de una aventurada reclamación de la corona. La bóveda de la capilla, magníficamente labrada, con sus múltiples formas circulares, se interpretó como una alabanza a un monarca ansioso por asegurarse un lugar para sí mismo en el orden cósmico e histórico. La bóveda nos dice más del patrono de la capilla y de su entorno social que de la personalidad del creador, sobre cuya identidad sólo podemos hacer conjeturas.

La bóveda de la capilla de Enrique VII representa la culminación de las bóvedas de abanico. En el año de 1540, su hijo, el rey Enrique VIII, recortó drásticamente el poderío y las riquezas de las órdenes monásticas, en cuyos edificios se habían levantado muchas bóvedas de abanico. La construcción de este tipo de bóvedas se paralizó durante casi cien años, hasta que fue revitalizada por la Iglesia de Inglaterra en la Universidad de Oxford, que era entonces un fuerte bastión del anglicanismo.

La construcción de las bóvedas de abanico nunca ha cesado por completo. Algunas catedrales modernas se han abovedado imitando los antiguos estilos. Los últimos ejemplos, no obstante, carecen del encanto y del interés de las bóvedas de abanicos construidas entre 1350 y 1540. El interés no sólo deriva de la belleza de las bóvedas en sí mismas, que es considerable, sino también del modo en que los constructores elaboraron las formas, quienes, aún actuando guiados solamente por impulsos estéticos y por resultados empíricos, se comportaron como expertos arquitectos.



8. TRACERIA DEL ABOVEDAMIENTO de la capilla de Enrique VII, el ejemplo más intrincado de los esquemas labrados sobre la superficie de una bóveda de abanico. La capilla se destinó a glorificar la dinastía Tudor; el labrado es un modo de expresar esa alabanza. La bóveda se construyó en su totalidad en estructura articulada, en la que la superficie de la misma la componían sillares cuidadosamente ajustados. Esa cualidad proporcionaba una mejor superficie en la que labrar luego los esmerados diseños.

Temas metamágicos

Las mil delicias del Lisp, lenguaje favorito en inteligencia artificial

Douglas R. Hofstadter

Con frecuencia, en artículos anteriores, he tocado el tema de la inteligencia artificial, que es la búsqueda de procedimientos para programar ordenadores de manera que su comportamiento pueda llegar a mostrar rasgos de flexibilidad, sentido común, creatividad, conciencia de sí y sentido del humor. Las indagaciones sobre inteligencia artificial comenzaron a ser tomadas seriamente hace más de veinte años, y desde entonces se han bifurcado muchas veces; hoy es terreno de rica y polimórfica investigación. Sólo en los Estados Unidos habrá un par de miles de personas profesionalmente implicadas en ese campo, y seguramente haya otras tantas en el resto del mundo. Aunque las divergencias de opinión son muy considerables entre los distintos investigadores, sí hay casi completa unanimidad en la elección del lenguaje de programación. Casi todos los intentos y ensayos en inteligencia artificial se llevan a cabo con auxilio de un lenguaje llamado Lisp, abreviatura de "list processing", "proceso de listas".

¿Por qué se realizan en Lisp casi todos los trabajos relativos a inteligencia artificial? Hay muchas razones, la mayoría de índole más o menos técnica, pero una de las más poderosas es verdaderamente sencilla: Lisp es un lenguaje claro y conciso. Todos los lenguajes de ordenador tienen algunos rasgos arbitrarios; de hecho, la mayoría están sobrecargados de ellos. Hay empero unos cuantos, como Lisp y Algol, que están construidos en torno a un núcleo de apariencia tan natural como una rama de las matemáticas. El núcleo del Lisp tiene una pureza cristalina que no sólo apela al sentido estético, sino que hace del mismo un lenguaje más flexible que la mayoría de los demás. A causa de su belleza y de la posición central que el Lisp ocupa en esta importante área de la ciencia moderna, dedicaré algunos artículos a presentar unas cuantas ideas que son fundamentales en este lenguaje.

Las raíces más profundas del Lisp calan, sobre todo, en la lógica matemática. Matemáticos pioneros, como Thoralf Skolem, Kurt Gödel y Alonzo Church aportaron a la lógica ideas seminales a lo largo de los años veinte y treinta, ideas que decenios más tarde se incorporarían al Lisp. La programación de ordenadores comenzó durante la década de 1940 a 1950, pero los llamados "lenguajes de alto nivel" no empezaron su andadura hasta el decenio siguiente. El primero de los lenguajes orientados al procesado de listas no fue el Lisp, sino el llamado IPL ("information-processing language") que fue desarrollado a mediados del decenio de 1950 por Herbert A. Simon, Allen Newell y J. C. Shaw. Durante los años 1956-58, John McCarthy, inspirándose en estas fuentes anteriores, puso a punto el elegante lenguaje algebraico de procesado de listas que él mismo bautizó "Lisp". Este lenguaje prendió rápidamente entre la joven pléyade que le rodeaba en el recién formado Proyecto de Inteligencia Artificial del Instituto de Tecnología de Massachusetts, fue materializado sobre el ordenador IBM 704, se propagó a otros grupos que trabajaban en inteligencia artificial y ha subsistido durante todos estos años. Aunque existen hoy muchos dialectos del lenguaje, todos comparten aquel mismo núcleo.

Pasemos a examinar la forma en que actúa Lisp. Una de las características más atractivas del Lisp es ser un lenguaje "interactivo", en contraste con la mayoría de los restantes lenguajes de alto nivel, que no lo son. Ser interactivo significa lo siguiente. Cuando se desea programar en Lisp, uno se acomoda frente a una terminal conectada a un ordenador que tenga en su memoria un sistema Lisp y teclea la palabra "lisp" (u otras a tal efecto). Inmediatamente veremos en nuestra pantalla una señal de "listo", un símbolo característico: una flecha o un asterisco. A mí me agrada imaginar que esta señal es un saludo que me dirige un "lisp-genio", quien, tras una profunda reverencia, nos dice: "Sus deseos son órdenes para mí. ¿Cuál es ahora su próxima voluntad?". El genio queda entonces a la espera de que usted se lo dactilografíe. Normalmente, al geniecillo se le conoce por "intérprete" de Lisp, y se encuentra dispuesto para hacer lo que se nos antoje..., pero es preciso poner el máximo cuidado para expresar con precisión nuestro deseo; de lo contrario las consecuencias pueden ser desastrosas. Vemos más abajo la señal de saludo de Franz Lisp (un dialecto particular), el signo con que el genio nos indica que está listo para ejecutar nuestras peticiones:

```
—>
```

El genio nos pregunta ahora qué deseos alberga nuestro corazón; teclearemos para empezar una expresión sencilla:

```
—> (suma 2 2)
```

y después pulsaremos la tecla de retorno del carro. Incluso los más profanos podrán prever que el Lisp-genio hará imprimir a continuación la cifra "4". Seguidamente, imprimirá también una señal de "a punto"; la pantalla mostrará por tanto:

```
—> (suma 2 2)
4
—>
```

El genio está listo ahora para ejecutar nuestra siguiente orden, o más cortésmente, nuestro deseo siguiente. La ejecución de un deseo expresado en forma de enunciado de Lisp se llama *evaluación* del enunciado. El breve intercambio recién expuesto entre humano y ordenador sirve para mostrar el comportamiento del intérprete Lisp: *lee* un enunciado, lo *evalúa*, *imprime* el valor apropiado e indica después estar listo para leer un nuevo enunciado. Por

esta razón, la actividad central del intérprete de Lisp suele conocerse por "bucle lectura-evaluación-impresión".

Es la existencia del Lisp-duende (el intérprete de Lisp) lo que hace que éste sea un lenguaje interactivo. Tan pronto como se dactilografía un deseo se obtiene realimentación inmediata. Y el procedimiento para conseguir que se realicen un conjunto de deseos es teclear uno, pedirle al genio que lo ejecute, manifestarle otro, volver a pedir al genio su realización, y así sucesivamente.

Por el contrario, en muchos lenguajes de programación "de alto nivel" (aunque no en todos, desde luego) es preciso escribir todo un programa, que consiste en un vasto número de deseos que han de ejecutarse en un orden especificado. Lo malo es que, por lo general, los deseos que podamos tener hacia el final dependen fuertemente de las consecuencias que hayan tenido deseos iniciales y, evidentemente, no tenemos posibilidad de irlos ensayando uno por uno. La ejecución de un tal programa puede, inútil es decirlo, conducir a resultados inesperados, a causa del gran número de deseos que tienen que armonizar perfectamente. De cometerse el más ínfimo error conceptual al diseñar su lista, lo más verosímil es que se produzca un embrollo total; en realidad es casi inevitable que así suceda. Hacer funcionar un programa de este tipo es como lanzar una nueva sonda espacial, todavía no probada: es imposible prever todo cuanto puede ir mal y, por tanto, lo único que podemos hacer es sentarnos a ver qué pasa, y confiar en que funcione. Si las cosas van mal se repasa y corrige aquello que parezca haber causado el fallo, y se efectúa un nuevo lanzamiento. Semejante proceder, desmañado, indirecto y caro contrasta fuertemente con el estilo directo, interactivo, a razón de un deseo por vez, que es característico del Lisp, y que permite el desarrollo, puesta a punto y depuración "incremental" de los programas. Y esta es otra razón de peso para explicar la popularidad del Lisp.

¿Qué clases de deseos podemos manifestarle al genio del Lisp, para que los evalúe, y qué tipos de cosas nos imprimirá como respuesta? Para empezar, podemos formularle expresiones aritméticas, aunque redactadas en forma bastante rara, como "(producto (suma 6 3)(diferencia 6 3))". La respuesta es 27, porque al evaluar "(suma 6 3)" da 9, "(diferencia 6 3)" da 3, y su producto es 27. Esta notación, donde cada una de las operaciones a ejecutar precede a sus operandos, fue ideada por el lógico polaco Jan Łukasiewicz antes de que existieran los ordenadores. Por desgracia para Łukasiewicz, su nombre presenta un aspecto demasiado formidable para la mayoría de los angloparlantes; este tipo de notación dio en ser llamada "notación polaca". He aquí un sencillo problema planteado en esta notación, donde usted debe hacer el papel del genio del Lisp:

```
—> (cociente
      (suma 21 13)
      (diferencia
        23
        (producto
          2
          (diferencia 7 (suma 2 2)))))
```

Tal vez haya usted observado que en los enunciados del Lisp se usan paréntesis. La profusión de paréntesis es una de las marcas de fábrica del Lisp. ¡No es excepcional encontrar expresiones donde hay que terminar cerrando una docena de paréntesis! Al principio muchos sienten escalofríos y, sin embargo, tan pronto uno se habitúa a sus característicos aspectos, las expresiones del Lisp se nos hacen intuitivas e incluso gratas de ver, especialmente cuando las imprimimos siguiendo un sistema de sangrados progresivos que refleje su estructura lógica (como se ha hecho en el ejemplo anterior).

El corazón del Lisp se encuentra en sus manipulables estructuras. Todos los programas en Lisp funcionan creando, modificando y destruyendo estructuras. Las estructuras son de dos tipos: atómicas y compuestas o, más brevemente, *átomos* y *listas*. Por consiguiente, todo objeto es, en Lisp, un átomo o una lista (pero no ambos), con la única excepción de un objeto especial llamado "nil", que es a un tiempo átomo y lista. Más adelante se hablará del nil. ¿Cómo son otros átomos típicos del Lisp? He aquí unos cuantos:

hidrógeno, helio, j-s-bach, 1729, pi, fu, bar, baz, botones-&-ojales.

Las listas son las estructuras de datos flexibles del Lisp. Una lista es en gran medida lo que dice ser: una colección de ciertas piezas en orden específico. Las piezas componentes de las listas se llaman *elementos* o *miembros* de ella. ¿Y qué objetos pueden ser elementos de una lista? No es sorprendente que las listas tengan átomos como elementos. Sin embargo, igual de fácilmente, las listas pueden tener como elementos a otras listas, las cuales pueden a su vez contener a otras, y así sucesiva y recursivamente. ¡Ay! Al decir esta palabra se me ha descargado sin querer la escopeta; menos mal que nadie ha resultado herido. Sin duda el lector ha comprendido a qué me refiero, y ello le servirá de preparación a una definición más técnica, que habrá de llegar algo más tarde.

Podemos reconocer las listas que nos presenta la pantalla de nuestra terminal por los paréntesis que las encierran. En Lisp, todo cuanto venga encerrado entre paréntesis concordantes es una lista. Por ejemplo "(zonk blis kopt (cronk flonk))" es una lista de cuatro elementos, el último de los cuales es a su vez una lista de dos elementos. Otra lista corta es "(suma 2 2)", ejemplo que muestra que los enunciados del Lisp también son listas. Y esto es importante, porque implica que el genio del Lisp, mediante manipulación de listas y átomos, puede construir efectivamente por sí mismo nuevos deseos. Por tanto, el objeto de un deseo puede ser la construcción –y evaluación subsiguiente– de un deseo enteramente nuevo.

Tenemos, por otra parte, la lista vacía, lista que no contiene elemento alguno. ¿Cómo se la denota? Podría pensarse que "()" bastaría. Y así es, en efecto, pero hay una segunda forma de indicar "lista vacía", y es escribir la palabra "nil". Ambas notaciones son sinónimas, aunque es más frecuente escribir "nil" que "()". La lista vacía, nil, es en Lisp concepto clave; en el universo de las listas hace igual papel que el cero en el universo de los números. Por usar otra metáfora, el nil es a modo de tierra donde enraízan y arraigan todas las demás estructuras. Empero, para comprender lo que esto significa, el paciente lector tendrá todavía que esperar un poco.

De las características de los átomos, a la que por lo común se le saca más partido es a la de tener (o poder asignárseles) valor. Ciertos átomos tienen valores permanentes; otros pueden tenerlos mudables. Como era de esperar, el valor del átomo "1729" es el número entero 1729, y éste valor es permanente. (Estoy distinguiendo aquí entre el átomo cuyo "nombre impreso" es la cadena de caracteres "1729", y la esencia platónica e imperecedera asociada con "ser el mínimo número expresable de dos maneras distintas como suma de dos cubos".) También el valor de nil es permanente, ¡y es nil! Hay

otro átomo, llamado "t", cuyo valor es permanente y también igual a sí mismo.

Aparte de t, nil y los átomos de nombre numérico, los átomos acostumbran a ser variables, lo que significa que es posible asignarles valores que más tarde podremos cambiar a voluntad. ¿Cómo se logra ello? Bueno, si queremos que al átomo "pie" le sea asignado el valor 4 podemos dactilografiarle al genio: "(hazq pie 4)". Podríamos, de la misma forma, indicarle "(hazq pie suma 2 2))", o "(hazq pie (suma 1 1 1 1))". En cualquiera de estos casos, tan pronto se pulse la tecla de retorno del carro el valor de "pie" se hará igual a 4, y así permanecerá en lo sucesivo, al menos, hasta que volvamos a efectuar una operación "hazq" sobre el átomo "pie".

Poca sería la capacidad de expresión del Lisp si los únicos valores que pudieran tomar los átomos fuesen valores numéricos. Afortunadamente, puede hacerse que el valor de un átomo sea cualquier objeto definido en Lisp: otro átomo, o una lista arbitraria. Por ejemplo, podríamos querer que el valor del átomo "pi" fuese una lista, como "(a b c)", o tal vez "(suma 2 2)", en lugar de ser el número 4. El valor correspondiente a un átomo suele llamarse *ligadura* del átomo, y se dice que el átomo está ligado a ese valor. Para asignar nuevas ligaduras a los átomos "pie" y "pi" volveremos a utilizar la operación "hazq". A fin de ilustrar cómo, presentamos aquí una breve conversación con el genio:

```
—> (hazq pie (suma 2 2))
4
—> (hazq pi '(suma 2 2))
(suma 2 2)
—>
```

Fijémonos en la inmensa diferencia entre los valores asignados a los átomos "pie" y "pi" a consecuencia de los dos deseos solicitados al genio, deseos que difieren únicamente en la ausencia o presencia de la mínima, pero crítica, simple comilla colocada ante la lista interior "(suma 2 2)". En el primer deseo, al no haber comilla, la lista "(suma 2 2)" interior tiene que ser *evaluada*. El resultado es 4, que es el nuevo valor asignado a la variable "pie". En el segundo deseo, por otra parte, en razón de la comilla a ella antepuesta, la lista "(suma 2 2)" no es ejecutada como instrucción, sino tratada como materia inerte, algo así como la carne del tablajero. Es carne "casi viva"; sin embargo, está muerta. Por consiguiente, el valor de pi en este segundo caso es la lista "(suma 2 2)", un fragmento en código Lisp. El siguiente diálogo con el genio nos confirma los valores de estos átomos:

```
—> pie
4
—> pi
(suma 2 2)
—> (eval pi)
4
—>
```

¿En qué ha consistido este último paso? He querido mostrar cómo se puede pedir al genio que *evalúe* el valor de una expresión, en lugar de limitarse a *imprimir* el valor de esa expresión. De ordinario efectúa automáticamente sólo un nivel de evaluación; escribiendo "eval" puede conseguirse una segunda etapa de evaluación (y evidentemente, reiterando eval una y otra vez puede continuarse tanto como se quiera). Esta característica resulta muchas veces de utilidad incalculable, pero es demasiado elevada para analizarla en este instante.

Exceptuada nil, todas las listas tienen al menos un elemento. El primer elemento de la lista se llama "car" de la lista. Así pues, car de "(eval pi)" es el átomo "eval". Los car de las listas "(suma 2 2)", "(hazq x 17)", "(eval pi)" y "(car pi)" son todos nombres de operaciones, o como suelen más comúnmente ser llamadas en Lisp, *funciones*. No es preciso que el car de una lista sea el nombre de una función; ni siquiera es preciso que sea un átomo. Por ejemplo, "((1) (2 2) (3 3 3))" es una lista perfectamente buena. Su car es la lista "(1)", cuyo car, a su vez, no es un nombre de función, sino una cifra.

Si tuviéramos que eliminar el car de una lista, ¿qué nos quedaría? Pues una lista más corta, que se llama "cdr" de la lista. Los términos "car" y "cdr" son reliquias de la primera versión del Lisp, instrumentalizada sobre un IBM 704; las letras de "car" y "cdr" denotan, respectivamente, "contents of the address part of register" y "contents of the decrement part of register", aproximadamente, "contenido de la parte de direccionamiento del registro" y "contenido de la parte decremental del registro", y aluden a características especiales del soporte físico, o *hardware*, de aquella máquina, que son, desde hace mucho, irrelevantes. El cdr de "(a b c)" es la lista "(b c)", cuyo cdr es "(c)", cuyo cdr es nil. Y nil no tiene cdr; al querer tomar el car o el cdr de nil se obliga al genio a emitir una tosecilla y un mensaje de error, lo mismo que al pedir una división por cero.

He aquí una tabla que muestra los car y cdr de unas cuantas listas, sin otra finalidad que hacernos ver que se trata de nociones definidas sin ambigüedad.

lista	*car*	*cdr*
((a) b (c))	(a)	(b (c))
(suma 2 2)	suma	(2 2)
((car x) (car y))	(car x)	((car y))
(nil nil nil nil)	nil	(nil nil nil)
(nil)	nil	nil
nil	ERROR	ERROR

De acuerdo con la terminología ordinaria, si car y cdr son llamados funciones, los objetos sobre los que actúen habrán de ser llamados *argumentos* suyos. Por tanto, en la instrucción "(suma pie 2)", "suma" es el nombre de la función y los argumentos son los átomos "pie" y "2". Al evaluar esta instrucción (y la mayoría de las demás) el genio calcula en primer lugar los valores de los argumentos, y les aplica después la función a los valores. Dado que el valor del átomo "pie" es 4 y que el valor del átomo "2" es 2, el genio nos retorna al átomo "6".

Supongamos que se tenga una lista, y que nos gustase disponer de una lista idéntica, aunque un elemento más larga. Por ejemplo, supongamos que el valor del átomo "x" sea "(talón, tobillo)" y que sea necesario crear una nueva lista, llamada "y", idéntica a la x aunque con un elemento más –"pie", pongamos por caso– al comienzo. Podemos usar entonces la función llamada "cons" (abreviatura de "construir") cuyo efecto es formar una nueva lista a partir de una lista antigua y de un car propuesto. He aquí la transcripción de un tal proceso:

```
—> (hazq x '(talón tobillo))
(talón tobillo)
—> (hazq y (cons 'pie x))
(pie talón tobillo)
—> x
(talón tobillo)
```

Vale la pena fijarse aquí en un par de cosas. Pedí la impresión del valor de x tras la primera operación de cons al objeto de que se pudiera ver que la lista x no quedaba modificada por tal operación. La función cons creó una nueva lista, e hizo que esta lista fuese el valor asignado a y, dejando x intacta. El otro hecho digno de atención es que hemos vuelto a utilizar la comilla antes del átomo "pie". ¿Qué hubiera sucedido de no

haberla utilizado? Veamos cual sería el resultado:

```
—> (hazq z (cons pie x))
(4 talón tobillo)
```

Recordemos que el átomo "pie" tiene asignado todavía el valor 4, y siempre que el genio encuentra en un deseo un átomo no afectado por una comilla toma siempre el valor perteneciente al átomo, y no el nombre del átomo (¿Siempre? Bien, casi siempre. Lo explicaré dentro de un instante. Mientras tanto, busque el lector una excepción; ya se ha tropezado con una.)

Ahora le propondré algunos ejercicios, algunos de cierta sutileza. ¡Atención a las comillas! Una última observación: he usado aquí la función "inversión", que genera una lista formada exactamente por los mismos elementos que su argumento, pero relacionados en orden inverso. Por ejemplo, al pedirle al genio "(inversa '((a b) (c d e)))", el resultado será "((c d e) (a b))". Damos seguidamente las frases dirigidas al genio en el diálogo:

```
—> (hazq w (cons pie '(cdr z)))
—> (hazq v (cons 'pie (cdr z)))
—> (hazq u (inversa v))
—> (cdr (cdr u))
—> (car (cdr u))
—> (cons (car (cdr u)) u)
—> u
—> (inversa
      '(cons (car u) (inversa (cdr u))))
—> (inversa
      (cons (car u) (inversa (cdr u))))
—> u
—> (cons 'tobillo
      (cons 'talón
         (cons 'pie nil)))
```

Las respuestas, tal como fueron impresas por el genio, son las siguientes:

```
(4 cdr z)
(pie talón tobillo)
(tobillo talón pie)
(pie)
talón
(talón tobillo talón pie)
(tobillo talón pie)
((inversa (cdr u)) (car u) cons)
(talón pie tobillo)
(tobillo talón pie)
(tobillo talón pie)
```

El último ejemplo presenta un proceso de aplicación repetitiva de la operación cons; este proceso suele llamarse "consar una lista" en la jerga del Lisp. Se parte de nil y se efectúan repetidamente operaciones de construcción. La situación recuerda la construcción de enteros positivos partiendo de cero y llevando a cabo iterativamente la operación "sucesor". Empero, mientras que en este segundo proceso hay en todo momento una única forma de llevar a cabo la operación sucesor, a partir de una lista hay infinitos objetos que "consar" (añadir) a ella, lo que engendra así un árbol inmensamente ramificado de listas que arrancan de la inicial, en contraste con la recta numérica, donde no puede haber ramificación alguna. Fue pensando en esta imagen de un árbol nacido del nil, y que contuviera todas las listas posibles, por lo que yo asimilé nil a la tierra en donde arraigan todas las posibles estructuras.

Como ya he mencionado, *no siempre* sustituye el genio por sus valores aquellos átomos que no vengan precedidos de comilla. Hay casos en que una función trata sus argumentos como si estuvieran precedidos por la comilla, a pesar de no estarlo. ¿Ha repasado usted los ejemplos? ¿Ha descubierto algún caso? No es difícil. La respuesta es la función "hazq". En particular, en una instrucción de tipo hazq, el primer átomo se toma tal cual esté, sin evaluarlo. De hecho, la "q" de "hazq" denota "quote" (cítese literalmente), queriendo decir que el primer argumento ha de ser tratado como si estuviera citado entre comillas. La situación se complica un poco al manejar la función "haz", parecida a la "hazq", pero que difiere de ésta en que *sí* efectúa la evaluación de su primer argumento. Por tanto, si el valor del átomo "x" es el átomo "k", al teclear "(haz x 7)", el átomo x no sufrirá modificación (su valor seguirá siendo el átomo "k") pero ahora el valor del átomo "k" habrá pasado a ser 7. Observe con atención:

```
—> (hazq a 'b)
—> (hazq b 'c)
—> (hazq c 'a)
—> (haz a c)
—> (haz c b)
```

Dígame usted ahora: ¿qué valores tienen los átomos "a", "b" y "c"? Doy aquí la respuesta, así que no caiga en la tentación de mirarla. Son respectivamente "a", "a" y "a". Tal vez le resulte desconcertante; quizá le tranquilice un poco saber que no es muy corriente usar en Lisp "haz", y que al usarlo, no es frecuente que aparezcan estas confusiones.

Desde el punto de vista psicológico, una de las grandes potencialidades de la programación reside en la posibilidad de definir operaciones compuestas, nuevas, a partir de otras anteriores, y de hacerlo una y otra vez, creando un vasto repertorio de operaciones cada vez más complejas. Es un proceso que nos trae reminiscencias de los primeros estadios de la evolución biológica, cuando a partir de moléculas sencillas fueron generándose evolutivamente otras más complejas, siguiendo a modo de una hélice ascendente de creatividad y complejidad. Nos hace pensar también en la Revolución Industrial, cuando la gente se valió de máquinas primitivas muy sencillas para construir máquinas más complejas, y seguidamente utilizó éstas para construir otras todavía más refinadas, según la misma hélice de complejidad y creatividad crecientes. En cada estadio, tanto de la evolución como de la revolución, los productos van haciéndose más y más versátiles e intrincados, más "inteligentes" si se quiere, pero también más vulnerables a sutiles averías y "pegas" de todo tipo.

Otro tanto sucede con la programación en Lisp, con la salvedad de que aquí las "moléculas" o "máquinas" son funciones del Lisp que se definen a partir de funciones previas, ya conocidas. Supongamos, por ejemplo, que conviniera disponer de una función que nos entregue siempre el último elemento de una lista, análoga a la "car", que nos proporciona el primero. Aunque el lenguaje Lisp no viene provisto de tal función, podemos fácilmente crearla nosotros mismos. ¿Ya se le ha ocurrido cómo? Para obtener el último elemento de una lista, que llamaremos "lysta", es suficiente invertir el orden de sus elementos y tomar car en la lista invertida: "(car (inversa lysta))". A fin de bautizar a esta operación con el nombre "rac" ("car" en orden inverso) nos valdremos de la función "def", como sigue:

```
—> (def rac (lambda (lysta)
      (car (inversa lysta))))
```

Al usar def en la forma anterior se crea la *definición* de una función. En ella, la palabra "lambda" seguida por "(lysta)" indica que la función que estamos definiendo tiene un solo parámetro o variable ficticia, que ha de ser llamada "lysta". (Podríamos haberle dado otro nombre cualquiera; lo que pasa es que a mí me gusta el átomo "lysta".) En general, la lista de parámetros ha de venir dada inmediatamente a continuación de la palabra "lambda". Una vez ejecutado este "deseo definitorio", nuestro genio comprende con pareja

claridad la función rac como antes la función car. Y ahora podemos hacer intervenir la propia función rac en la definición de nuevas funciones. La cosa crece tan exponencialmente como una bola de nieve, y pronto nos sentimos abrumados por el poder puesto en nuestras manos.

Veamos un ejemplo sencillo. Imaginémosnos una situación que exija manejar muchas listas largas, y que se sepa que frecuentemente podrá resultarnos útil disponer para cada una de ellas de otra muy breve que contenga justamente su car y su rac. Podemos definir entonces una función uniparamétrica que nos lo resuelva:

```
(def resumen-tipo-readers-digest
   (lambda (listalarga)
   (cons (car listalarga)
      (cons (rac listalarga) nil))))
```

Así pues, al aplicar la función "resumen-tipo-readers-digest" al texto íntegro de *El Quijote* de Miguel de Cervantes (tratándolo meramente como una larga lista de palabras) obtendremos la lista abreviada "(En Quijote)". Desafortunadamente, al aplicar de nuevo la función de condensación a esta lista ya no se consigue simplificarla más.

Sería grato y útil poder disponer de una función inversa de la llamada "resumen-tipo-readers-digest" (a la que podríamos titular "recervantes") tal que dadas dos palabras cualesquiera construyese una novela de Cervantes que comenzase y terminase precisamente por ellas. De lograrla, la ejecución del enunciado "(recervantes 'Paseándose 'soldado)" tendría por efecto que el genio de Lisp redactase, partiendo de cero, la totalidad de su novela *El licenciado Vidriera*. La definición de la función se deja, como ejercicio, al cuidado del lector.

Un objetivo que algunos han encontrado deseable y factible en Lisp y en otros lenguajes de programación similares es (1) hacer que todo enunciado nos retorne un *valor* y (2) hacer que sea por intermedio de este valor, y *únicamente* a través de él como puedan producirse los efectos del enunciado. La idea que inspira el objetivo (1) es que los valores se vayan entregando y vayan "emergiendo" desde las llamadas a las funciones más internas hacia las más periféricas, hasta el valor final, que nos es retornado. En (2) la idea es que durante todo el proceso de emersión ningún átomo sufra modificación de su valor (a menos que el átomo sea una variable ficticia). En todos los dialectos de Lisp que me son conocidos (1) es verdadero, pero (2) no necesariamente lo es.

Así pues, si x está ligado con "(a b c d e)" y pedimos "(car (cdr (inversa x)))", lo primero que sucede es que se calcula el valor de "(inversa x)"; seguidamente, este valor se entrega a la función cdr, la cual determina el cdr de esa lista; finalmente, esta lista abreviada le es suministrada como dato a la función car, la cual extrae de ella un elemento, el átomo "d", que nos entrega de retorno, como salida. En el ínterin, el átomo "x" no ha sufrido daño alguno, y sigue ligado a "(a b c d e)".

Podría pensarse que expresiones como "(inversa x)" cambiarían el valor de x en el proceso de inversión, lo mismo que al obedecer nosotros la orden "vuélvete el suéter del revés" el suéter sufrirá las consecuencias. En realidad, en la ejecución del deseo "(inversa x)", la lista x no sufre mayor daño del que pueda sufrir el átomo 2 al efectuar "(suma 2 2)". En lugar de modificar x, la ejecución de "(inversa x)" trae al mundo una lista *nueva*, sin nombre, idéntica a la x excepto por el orden de sus elementos. Y esa lista es el valor del enunciado; es lo que retorna el enunciado. El valor de x, sin embargo, permanece intacto. Análogamente, al evaluar "(cons 5 pi)" la lista denominada "pi" no cambia en lo más mínimo; lo único que hace es dar como *retorno* una lista nueva cuyo car es 5 y cuyo cdr es el valor que pi pudiera tener.

Es preciso contrastar esta conducta con la de funciones que provocan "efectos secundarios". Por lo común tales efectos secundarios vienen en la forma de cambios en las ligaduras de las variables, si bien hay otras posibilidades, como por ejemplo, que tengan lugar operaciones de entrada o salida. Una de las instrucciones "perniciosas" típicas es hazq; por ello, los defensores de la escuela de programación "aplicativa" –que mantiene que ninguna instrucción debe provocar efectos secundarios de ningún tipo– quedan profundamente trastornados con sólo mentarles hazq. Según los aplicativistas, todos los resultados que se produzcan deberían ser consecuencia estricta de la forma en que las funciones computan sus valores y se los van traspasando de unas a otras.

Las únicas ligaduras que merecen el beneplácito de los puristas del estilo aplicativo son las ligaduras transitorias "de tipo lambda", que surgen al aplicar funciones a los argumentos de éstas. Cada vez que se invoca una función, las variables ficticias que en ella intervienen asumen ligaduras "lambda". Tales ligaduras son idénticas a las producidas por una instrucción hazq, pero su duración es efímera: tan pronto queda concluido el cómputo del valor de la función, las ligaduras "lambda" se desvanecen, esfumándose sin dejar rastro. Por ejemplo, al calcular "(rac '(a b c))" se establece una ligadura "lambda" entre la variable ficticia "lysta" y la lista "(a b c)"; pero tan pronto la respuesta "c" queda en manos de la persona o la función que solicitó el rac, el valor del átomo "lysta" utilizado para calcularlo queda olvidado por completo. Si pedimos al ordenador el valor de "lysta", el intérprete de Lisp nos dirá que el átomo "lysta" es un átomo "desligado". Los programadores de la escuela aplicativa prefieren con mucho las ligaduras "lambda" a las de tipo hazq ordinarias.

Personalmente, yo no soy un obseso con respecto a que deban evitarse funciones como hazq, y otras, que puedan provocar efectos secundarios. Aunque encuentro elegante el estilo aplicativo, lo tengo por poco práctico para redactar programas largos en inteligencia artificial. Por consiguiente, no abogaré aquí por el estilo aplicativo, aunque estoy dispuesto a adaptarme a él siempre que sea posible. Estrictamente hablando, en programación aplicativa ni siquiera sería lícito definir nuevas funciones, porque los enunciados de tipo "def" provocan cambios permanentes en la memoria del genio, a saber, almacenan permanentemente en ella la definición de la función. Por tanto, desde el enfoque aplicativo radical, las definiciones de las funciones, lo mismo que las ligaduras de las variables, se crearían sólo temporalmente, y tan pronto fuesen utilizadas habrían de ser desechadas. Sería ésta una forma ideal y extrema de "aplicativismo".

Para edificación del lector, he aquí otras cuantas definiciones de funciones sencillas:

```
—> (def rdc
       (lambda (lysta)
       (inversa (cdr (inversa lysta)))))
—> (def snoc
     (lambda (x lysta)
        (inversa
           (cons x (inversa lysta)))))
— > (def doble (lambda (n) (suma n n)))
```

Las funciones rdc y snoc son análogas a las cdr y cons, tomadas en sentido inverso. Por tanto, rdc de "(a b c d e)" es "(a b c d)", y si solicitamos "(snoc 5 '(1

2 3 4))", la respuesta que obtendremos será "(1 2 3 4 5)".

Todo esto ha sido, hasta ahora, levemente interesante, nada más. Si queremos ver al genio hacer algo realmente sorprendente tendremos que dejarle tomar algunas decisiones basadas en acontecimientos acaecidos por el camino, y que acostumbran a llamarse deseos *condicionales*. Un ejemplo típico sería el siguiente:

```
—> (cond ((eq x 1) 'tierra)
         ((eq x 2) 'mar))
```

El valor retornado por este enunciado será el átomo "tierra" si x tiene el valor 1, y el átomo "mar" si x tiene el valor 2. En los restantes casos, el valor retornado será nil (así ocurrirá, por ejemplo, si x vale 5). El átomo "eq" (iniciales de "equal", "igual") es el nombre de una función muy utilizada en Lisp, que retorna el átomo "t" (inicial de "true", "verdadero") si sus dos argumentos tienen igual valor, y nil (correspondiente a "no" o "falso") si lo tienen distinto.

Un enunciado condicional es una lista cuyo car es el nombre funcional "cond", seguido de cualquier número de *cláusulas condicionales*, cada una de las cuales es una lista de dos elementos. El primer elemento de la cláusula es la "condición", el segundo elemento es el "resultado". Las condiciones de las cláusulas se van examinando, una a una, por el Lisp-genio, en el mismo orden en que vengan dadas; tan pronto como el genio da con un cláusula cuya condición es "verdadera" (lo que significa que el retorno de la condición es distinto de nil) comienza a calcular el *resultado* de la cláusula, cuyo valor será el retornado correspondientemente al enunciado completo. ¡Ninguna de las cláusulas adicionales que puedan figurar recibe del genio tan siquiera una ojeada! Quizá le parezca esto más complejo de lo que debiera; no hay realmente complicación alguna: el genio va buscando hasta encontrar la primera condición satisfecha, y retorna como valor de todo el enunciado el resultado correspondiente a dicha condición.

Conviene con frecuencia situar al final del enunciado una cláusula "centinela" cuya condición se cumpla con certeza; de esta manera, aunque fallen todas las demás condiciones, el retorno será el resultado que la acompañe, nunca un nil. Crear una condición cuyo valor sea no-nil es tan sencillo como comerse un pastel; basta, por ejemplo, tomar como condición el átomo "t". Así hemos hecho aquí:

```
—> (cond ((eq x 1) 'tierra)
         ((eq x 2) 'mar)
         (t 'aire))
```

Según sea el valor de x obtendremos como valor del enunciado, ya "tierra", ya "mar", o bien "aire". Lo que nunca obtendremos, en cambio, será un nil. He aquí unos cuantos enunciados más de tipo condicional, para que usted haga en ellos el papel del genio:

```
—> (cond ((eq (eval pi) pie)
          (eval (snoc pie pi)))
        (t (eval (snoc (rac pi) pi))))
—> (cond ((eq 2 2) 2) ((eq 3 3 ) 3))
—> (cond (nil 'no-no-no)
          ((eq '(car nil) '(cdr nil))
          'hmmm)
          (t 'si-si-si))
```

Las respuestas son 8, 2 y "si-si-si". ¿Se ha fijado usted en que "(car nil)" y "(cdr nil)" estaban antecomilladas?

Terminaré esta parte del artículo mostrando una familia de definiciones de funciones, que se ajustan a una pauta tan evidente que casi podríamos pensar que el genio del Lisp "les cogería el tranquillo" apenas vistas unas cuantas. Desdichadamente, los Lisp-genios son de una cerrazón exasperante (o al menos juegan a serlo), e incapaces de hacer inferencia alguna que no se les haya explicitado por completo. Veamos primero la familia en cuestión:

```
—> (def cuadrado
      (lambda (k) (producto k k)))
—> (def cubo
      (lambda (k) (producto k (cuadrado k))))
—> (def 4.ª-potencia
      (lambda (k) (producto k (cubo k))))
—> (def 5.ª-potencia
      (lambda (k)
       (producto k (4.ª-potencia k))))
—> (def 6.ª-potencia
  (lambda (k)
:   (producto k (5.ª-potencia k))))
:
:
```

He aquí la cuestión que le planteo: ¿sabrá usted inventar la definición de una función *bi*-paramétrica que condense "de una tacada" esta familia infinita? Más concretamente, la cuestión es ésta: ¿cómo deberíamos proceder para definir una función de dos parámetros, llamada "potenciación", o más brevemente, "pot", tal que, por ejemplo, "(pot 9 3)" produzca 729 al ser evaluada, y "(pot 7 4)" retorne 2401? Ya he dado en esta sección todo el instrumental necesario para construirla; le bastará, pues, emplear un poco su imaginación.

He pensado en terminar el artículo con una primicia periodística: el descubrimiento del cuerpoespín glazunkiano, efectuado hace pocas fechas, y así llamado por ser su hábitat exclusivo la isla de Glazunkia (que no obstante encontrarse a tiro de piedra de la costa de Burronymede es reclamada como propia por Bitbo del Norte). ¿Y qué es un cuerpoespín, se preguntará el lector? Pues una extraña variedad de puercoespín, cuyas púas –siempre nueve en Glazunkia Exterior y siete en Glazunkia Interior, por razones que se ignoran todavía– son cuerpoespines más pequeños. ¡Ya estamos! Parece que acabamos de tropezarnos con una regresión infinita. Pero no. Lo que sucede es que se me ha olvidado decirles que hay un tipo de cuerpoespín más pequeño, el de tamaño cero, que curiosamente carece totalmente de púas. Es una verdadera suerte (o mala suerte, según su punto de vista) que este hecho ponga término a la regresión infinita que parecía venírsenos encima.

Los estudiosos de la zoología pueden tener interés en el dato de que los cuerpoespines de tamaño cinco tienen espinas que son cuerpoespines de tamaño cuatro, y así en sucesión decreciente. Los antropólogos pueden encontrar intrigante que los residentes de Glazunkia (tanto Interior como Exterior) utilicen el hocico (¡sí, el hocico!) del cuerpoespín de tamaño cero como unidad de trueque, cosa que resulta bien chocante a nuestra mentalidad, aunque, ¿quienes somos nosotros para poner en tela de juicio las tradiciones ancestrales de los glazunkianos? Puesto que un cuerpoespín medianejo –de tamaño tres o cuatro, por ejemplo– tiene una multitud de estos hociquillos, poseerlos resulta de la máxima importancia. El valor de trueque de un cuerpoespín se llama "capacidad adquisitiva", "poder de compra", o simplemente, "poder". Por ejemplo, uno de tamaño dos capturado en Glazunkia Interior tiene casi el doble de poder adquisitivo que otro de igual tamaño atrapado en Glazunkia Exterior. ¿O es al revés?

A fin de cuentas, ¿por qué les explico todo esto? Bueno, pensé que les gustaría saberlo. Además, podría ocurrir que alguno de ustedes acabase visitando Glazunkia (Interior o Exterior) uno de estos días, y entonces todo esto les resultaría de la mayor utilidad.

Taller y laboratorio

Osciladores térmicos: sistemas que zumban, se balancean o ululan a instancias del calor

Jearl Walker

Los osciladores térmicos entran en los programas de física dentro del apartado dedicado a los sistemas oscilantes alineales y automantenidos. En ellos, una parte del sistema vibra ininterrumpidamente porque recibe calor. Voy a ocuparme aquí de cuatro de esos ingenios; dos son nuevos y los otros dos ya conocidos en el siglo XIX. Los cuatro se fundamentan en una redistribución periódica del calor, pero son completamente distintos entre sí.

Pierre Welander, de la Universidad de Washington, ha ideado un original oscilador térmico que funciona a base de agua. Consiste en unos tubos comunicantes, cuyos dos extremos están conectados al fondo de sendos recipientes anchos. Los tubos y aproximadamente la mitad de cada recipiente están llenos de agua. Los recipientes quedan abiertos por arriba. Cuando el codo de unión de los tubos recibe calor, mientras que los extremos superiores de éstos se refrigeran. El nivel del agua contenida en el artefacto comienza a oscilar entre uno y otro tubo, con un período comprendido entre varios minutos y varias horas.

Supongamos que el sistema se encuentra inicialmente en equilibrio, conteniendo ambos tubos la misma cantidad de agua con iguales distribuciones de temperaturas. Fijémonos en el diminuto volumen de agua encerrado en una rebanada muy fina situada en la sección transversal del centro del codo de unión. Este volumen estará comprimido en ambas caras en virtud del peso del agua existente en cada tubo.

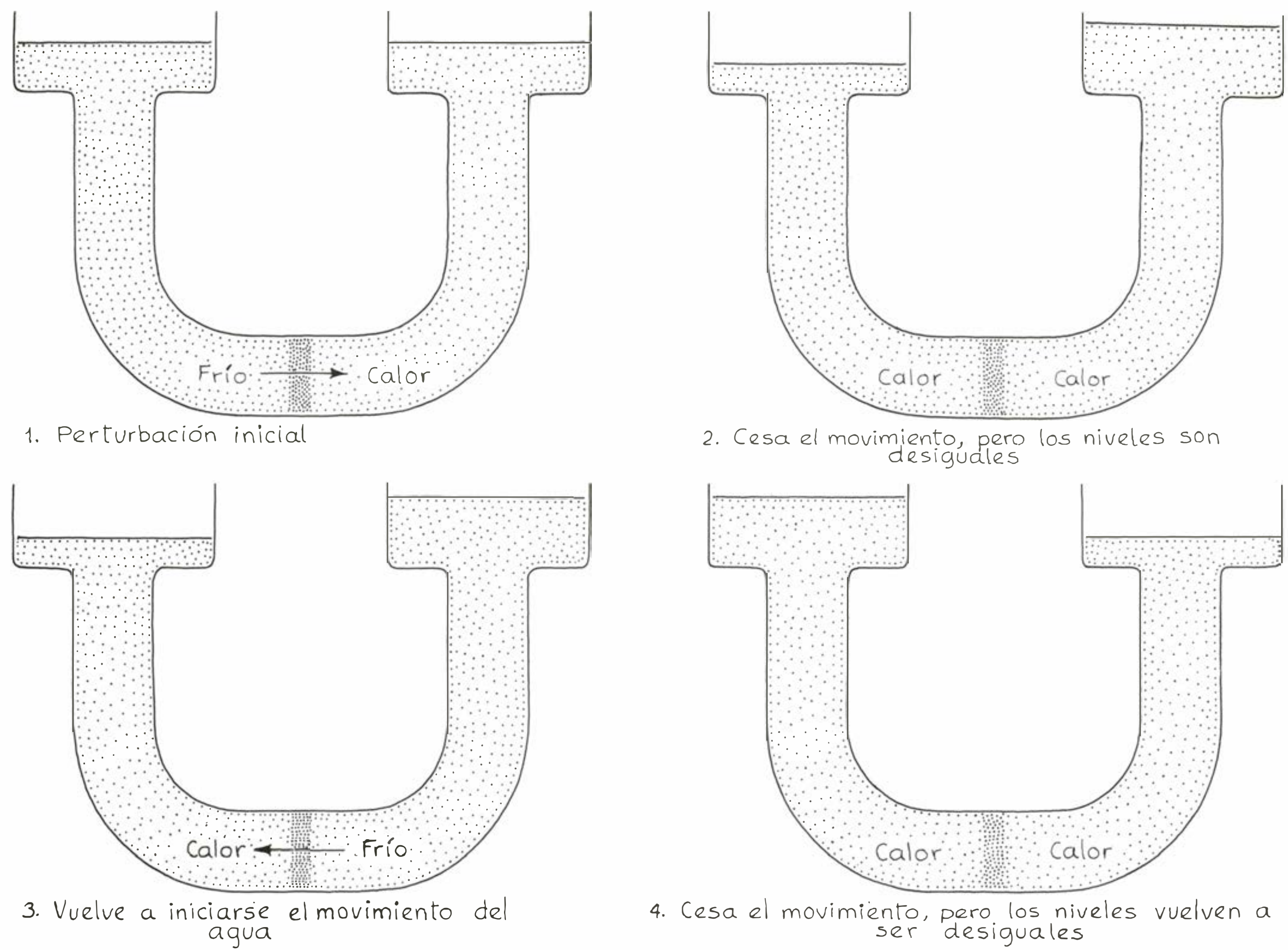
El peso de una columna de agua lo determinan dos cosas: la altura de la columna y la densidad del agua. La densidad depende de la temperatura: el agua más fría es más densa que la más caliente. Como, al principio, los dos tubos tienen el mismo nivel de agua y la misma distribución de temperaturas, el volumen de agua central está comprimido por igual en ambas caras y, por tanto, se encuentra inmóvil y toda esa masa de agua está en equilibrio.

Supongamos que una perturbación procedente del entorno provoque el desplazamiento de derecha a izquierda de una pequeña cantidad de agua. Este movimiento romperá la simetría del sistema: ahora ambos tubos diferirán en la altura de sus niveles de agua y en las distribuciones de temperaturas. En efecto, con ese movimiento, el nivel del agua en el tubo izquierdo descenderá y bajará agua fría, mientras que subirá el nivel en el tubo derecho y a la parte derecha del codo de unión afluirá agua más caliente.

Si los recipientes conectados a los extremos de los tubos son anchos, las variaciones de los niveles de agua serán inicialmente pequeñas. Mayores consecuencias acarrea la ruptura de la simetría en la distribución de temperatura. Como el agua más fría arribada a la parte izquierda del codo es más densa que la desplazada hacia la parte derecha, el volumen de agua central ya no estará sometido a la misma presión en ambas caras, sino que será mayor la que sufra la cara izquierda. Desigualdad de presiones que tenderá a impulsar más agua hacia la derecha.

Aunque la variación inicial de los niveles de agua sea pequeña, el paso continuo de un lado a otro acrecienta la diferencia hasta que ésta deviene importante. Con ello, y a consecuencia de la diferencia de niveles, el tubo derecho acabará ejerciendo sobre el volumen central una presión mayor que el tubo izquierdo. La corriente de izquierda a derecha remite y, muy pronto, cesa. Se trata de un proceso gradual; el calentamiento y la refrigeración tienen ocasión de restablecer la simetría de temperaturas. En el instante en que cesa la corriente, la distribución de temperaturas del sistema vuelve a ser aproximadamente simétrica; pero entonces la altura del agua será mayor en el tubo derecho que en el izquierdo y el agua comenzará a regresar al tubo izquierdo, reanudándose el ciclo en sentido contrario. El sistema permanece en oscilación.

1. Condiciones físicas del oscilador térmico ideado por Pierre Welander



2. Esquema de la corriente en un oscilador fluido

Esta corriente oscilante proseguirá indefinidamente a causa del calentamiento y refrigeración de los tubos. Si se deja que éstos alcancen una temperatura uniforme, la oscilación se amortiguará sin tardanza hasta un estado estable con iguales niveles de agua en los dos tubos. Pero mientras continúen el calentamiento y la refrigeración, persistirá la oscilación. Y ello porque cada vez que el sistema atraviese el estado en que ambos niveles son iguales, el desequilibrio de densidades impulsará al agua hasta hacerlos desiguales, manteniendo la oscilación.

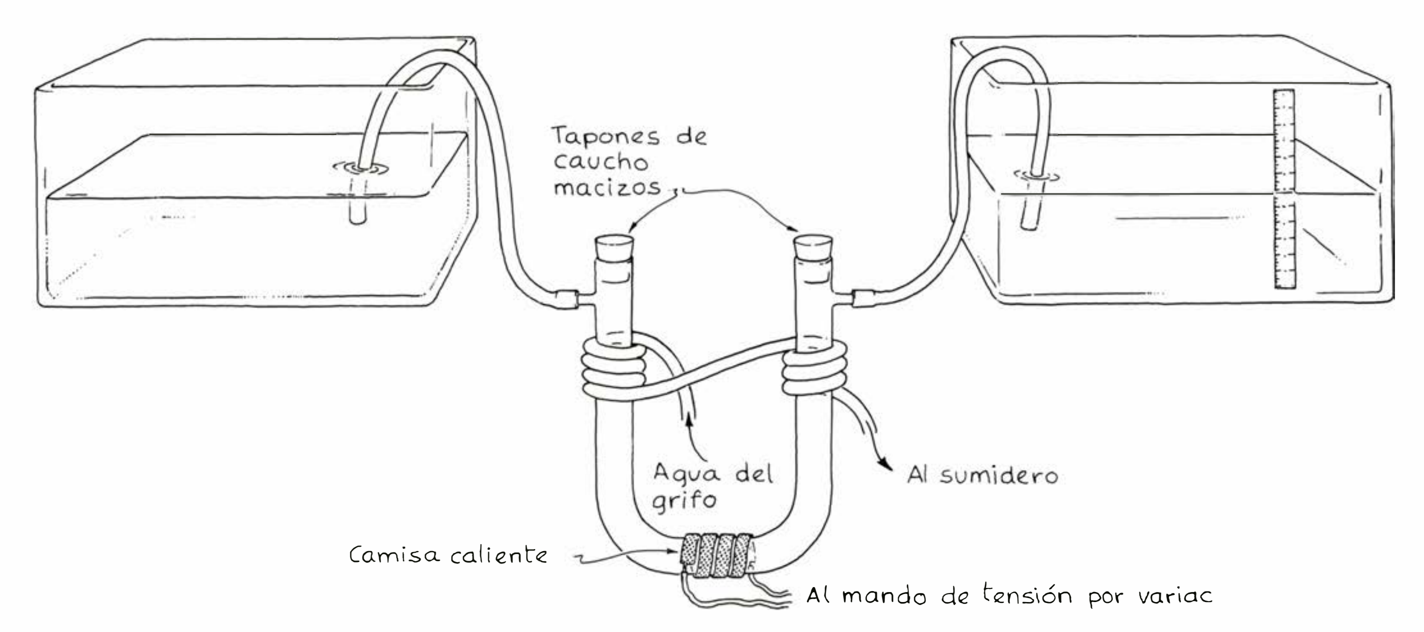
Para que el sistema entre en oscilación tiene que reunir ciertas características geométricas. La superficie superior de los recipientes será mayor que la sección transversal de los tubos en una cuantía que es función de la distribución de temperaturas en el agua. Así, el cociente de la sección de cada recipiente entre la sección del tubo debe ser mayor que 5000 dividido por la diferencia de temperaturas máximas que se establezca en el agua merced al calentamiento y refrigeración.

Por ejemplo, si el montaje produce una diferencia de temperaturas de 50 grados Celsius, la superficie de cada depósito será al menos 100 veces mayor que la sección de los tubos. Una diferencia de temperaturas menor exigiría un cociente entre superficies aún mayor. Por supuesto, si los depósitos son desmesuradamente grandes, quizá cueste percibir la oscilación de los niveles de agua.

Esta condición que deben cumplir las dimensiones relativas tiene su origen en la interacción de las presiones sobre el volumen central considerado del codo de unión. Supongamos que una perturbación exterior inicial impulsa una pequeña cantidad de agua hacia la derecha. La asimetría de temperaturas consiguiente tenderá a impulsar más agua hacia la derecha, a lo que se opone la diferencia entre los niveles de agua en ambos lados. Entonces, si la superficie de los depósitos es demasiado pequeña, el paso del agua creará inmediatamente una diferencia de niveles más que suficiente para detener la corriente producida por la asimetría de temperaturas. El equilibrio se restablecerá rápidamente.

Para que se inicien las oscilaciones, las áreas superficiales de la parte superior de los depósitos habrán de alcanzar cierto tamaño. De ese modo, la corriente provocada por la asimetría de temperaturas no producirá inmediatamente una gran diferencia de niveles. Así, la corriente de agua dispondrá de tiempo para establecerse. El período de las oscilaciones lo fijan las áreas relativas de los depósitos y los tubos, la longitud de éstos, la viscosidad del agua y la aceleración de la gravedad.

Puede construirse un artefacto semejante al de Welander soldando unos depósitos de vidrio a los extremos de unos tubos comunicantes. Como éstos deben ser de pyrex para que no salten en añicos a consecuencia del calentamiento y enfriamiento, la soldadura de los depósitos me pareció excesivamente dificultosa. Tampoco vi fácil montar un sistema tan delicado que permaneciera vertical. Busqué unos tubos comunicantes provistos de ramas laterales junto a los extremos; las conecté a unos recipientes de plástico mediante trozos de manguera de caucho. Los recipientes eran iguales: la sección transversal



3. Dispositivo oscilador similar al de Welander

superior de los dos era unas 200 veces mayor que la sección transversal de los tubos.

Con cinta adhesiva sujeté una regla transparente a la pared interior de uno de los recipientes, de modo que pudiera controlar el nivel del agua. Llené de agua los tubos y, hasta la mitad aproximadamente, los recipientes. En torno al fondo del codo de unión arrollé una camisa calefactora de las que utiliza cualquier laboratorio de química. Contiene la camisa un conductor eléctrico de gran resistencia que genera calor. La conecté a un mando de tensión con autotransformador de relación variable (Variac), y así poder ajustar la cantidad de calor suministrada a los tubos. (Me esmeré muchísimo para asegurarme de que ninguna infiltración de agua repentina alcanzara los conductores.)

Como elemento enfriador empleé manguera de caucho de pared fina por cuyo interior circularía agua del grifo. La arrollé en varias vueltas alrededor de los extremos de los tubos. Aunque con este procedimiento conseguía enfriar, desde luego, el agua situada en la parte superior de los tubos, llegué a la conclusión de que necesitaba un método más eficaz, ya que la conducción de calor a través de la pared de vidrio de los tubos y el caucho de la manguera me pareció excesivamente lenta.

No disponía de método para medir la distribución de temperaturas en el agua contenida en los tubos. Tras experimentar ajustando el Variac de varios modos, logré mi propósito de observar pequeñas oscilaciones de los niveles del agua de los recipientes. El correspondiente a la regla se elevaba y descendía a razón de un milímetro aproximadamente cada cinco minutos. Estas oscilaciones no eran amplias ni siempre regularmente espaciadas; pero eran del tipo de las observadas por Welander: desaparecían al suprimir los dispositivos de calentamiento y enfriamiento.

Welander me señaló un indicador más sensible del nivel de agua. De un brazo corto de alambre se cuelga un tapón de corcho desde el borde del recipiente, de modo que flote en el agua. En el corcho se instala una pajita vertical. Cuando varía el nivel del agua, el corcho flotante rota en torno al brazo de alambre y el extremo superior de la pajita amplifica el movimiento.

Welander ha descrito asimismo un oscilador térmico de otro tipo. Dentro de un recipiente con agua se forma periódicamente una capa de hielo. Este experimento lo realizó con unos 50 litros de agua en una tinaja de plexiglás, aislada térmicamente, cuyo fondo era un cuadrado de 50 cm de lado. Junto al fondo dispuso un elemento calefactor, capaz para unos 100 watt. Este calor lo distribuían unos agitadores magnéticos que mezclaban continuamente el agua. El dispositivo lo almacenó en un compartimento cuyo aire se mantenía a 20 grados Celsius. Además, el aire situado encima del agua estaba asimismo sometido a una agitación continua.

Welander registraba la temperatura del agua a media altura del recipiente. Así, se dedicó a observar la formación de hielo: la superficie del agua se helaba y fundía periódicamente cuando la temperatura del líquido variaba aproximadamente en unos 10 grados. Durante la fase de la oscilación en que la superficie se helaba, la temperatura del agua crecía hasta unos 10 grados Celsius para disminuir casi hasta el punto de solidificación durante la fase sin hielo. El período de oscilación rondaba las 50 horas.

Cualquiera puede observar oscilaciones de solidificación y fusión similares en su propia casa. Welander sugiere emplear un termo de boca ancha. Este se llena de agua hasta arriba, se introduce un serpentín de calefacción dentro del agua y se guarda el conjunto en un congelador. Entonces, se formará y deshará hielo periódicamente, aunque las oscilaciones no serán tan uniformes como las observadas por Welander en su dispositivo de mayor tamaño.

He repetido el procedimiento, con un ligero cambio. Por serpentín de calefacción empleé el calefactor de inmersión que me sirve para preparar el agua del café instantáneo. Lo suspendí en la posición correcta dentro del termo valiéndome de varias tiras de cinta adhesiva reforzada. Pero la envuelta metálica del calefactor sobresalía por la superficie del agua, con lo que el calor que desprendiese interferiría con la formación periódica de hielo.

Enchufé entonces el calefactor a un Variac para tantear la velocidad a la que se cedía calor al agua. La junta de caucho que reviste los bordes de la puerta de mi congelador apretaba lo suficiente para permitirme cerrarla por completo sobre el cordón de alimentación. Si así no hubiera sido, ya me encontraba preparado para rellenar el borde con tela, que bloqueara el paso de aire ambiental. Puse el mando del congelador al máximo. Aquí también tomé mis precauciones para que el agua no se derramará sobre la alimentación eléctrica.

Las oscilaciones en la formación de hielo se deben a la mala transmisión del calor a través del hielo (en comparación con la transmisión a través del agua agitada mecánicamente o bien por convección). El calor que cede al sistema la resistencia eléctrica sumergida debe propagarse hacia arriba, porque los costados del recipiente están aislados. Además, la superficie del agua se enfría merced a un aire que se encuentra a una temperatura inferior a su punto de solificación.

El sistema puede evolucionar hacia dos estados estacionarios posibles. En uno de ellos, la superficie del agua permanece líquida porque el aire no se encuentra especialmente frío y el calor que suministra la resistencia impide siempre la congelación. Como el calor se propaga hacia arriba por conducción y convección dentro del agua, la transmisión de calor, suficientemente rápida, impide la solidificación.

En el otro estado estacionario, la parte superior del sistema se halla siempre solidificada porque el aire está muy frío y la velocidad de aporte de calor no es elevada. El calor cedido por la resistencia es transportado hasta la capa de hielo por conducción y convección. En ese lugar, atraviesa el hielo lentamente y las corrientes convectivas en el seno del aire lo arrastran. En este caso, el estado del sistema es estacionario porque, aunque la conducción a través del hielo sea lenta, no lo es hasta el punto de impedir que la temperatura del agua subyacente se eleva hasta fundir el hielo.

Las oscilaciones pueden desencadenarse cuando el aire esté muy frío y, alta, la velocidad de aportación de calor por parte de la resistencia. Supongamos que el agua no esté solidificada. Entonces el aire robará calor a la superficie con rapidez suficiente para que comience la solidificación. Una vez formada la capa de hielo, disminuye la eliminación de calor del sistema, pues el hielo no es un transmisor térmico de tan buena calidad como el agua líquida agitada. Por tanto, la temperatura del agua situada bajo el hielo comienza a crecer, hasta que se calienta lo suficiente para fundir el hielo.

Cuando este último desaparece, se acrecienta la velocidad de transferencia de calor y disminuye la temperatura del agua. Posteriormente, la temperatura de la superficie retornará al punto de solidificación y se formará una nueva capa de hielo.

En 1805, un inspector descubrió otro tipo de oscilador térmico en una fundería de Sajonia. Cuando una barra de metal caliente se deposita sobre un bloque de otro metal, la primera puede oscilar ruidosa y enérgicamente. Este mismo oscilador lo redescubrió, 25 años después, Arthur Trevelyan, razón por la cual se conoce desde entonces como balancín de Trevelyan.

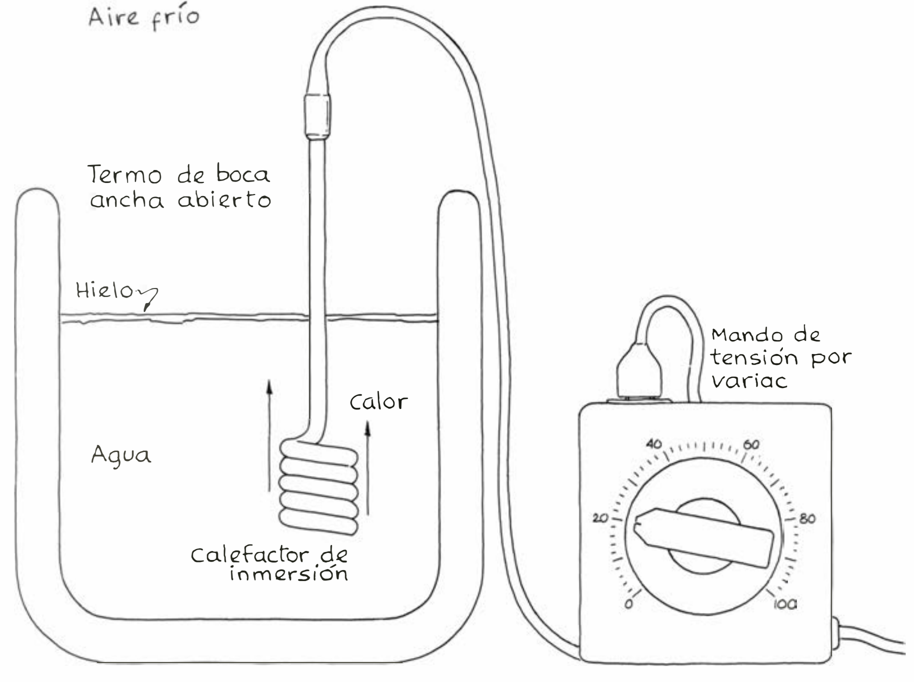
He visto dos versiones de este tipo de oscilador. En ambas importa acertar en la elección de los metales. En una se montan dos placas de plomo en un tornillo de banco y se afilan sus cantos superiores dejando una pequeña separación entre ambos. Luego se coloca una barra de latón en equilibrio entre una y otra placa. Entonces, cuando la barra se calienta, comienza a mecerse, de modo tal que cuando da contra el canto de una placa y, después, contra el de la otra, organiza un golpeteo casi rítmico.

Para la segunda versión se necesita un prisma de latón de sección triangular pero con dos bordes afilados en uno de los vértices del triángulo. Estos descansan perpendicularmente sobre la arista de un bloque de plomo. Solidaria por un extremo al trozo de latón y paralelamente a los dos bordes afilados se extiende una varilla de latón provista de un peso en el otro extremo, junto al cual la varilla descansa libremente sobre un soporte. Cuando la varilla se calienta, los bordes afilados del trozo de latón se mecen en torno al eje de la varilla, produciendo un sonido de frecuencia aproximadamente constante en su percusión contra el bloque de plomo.

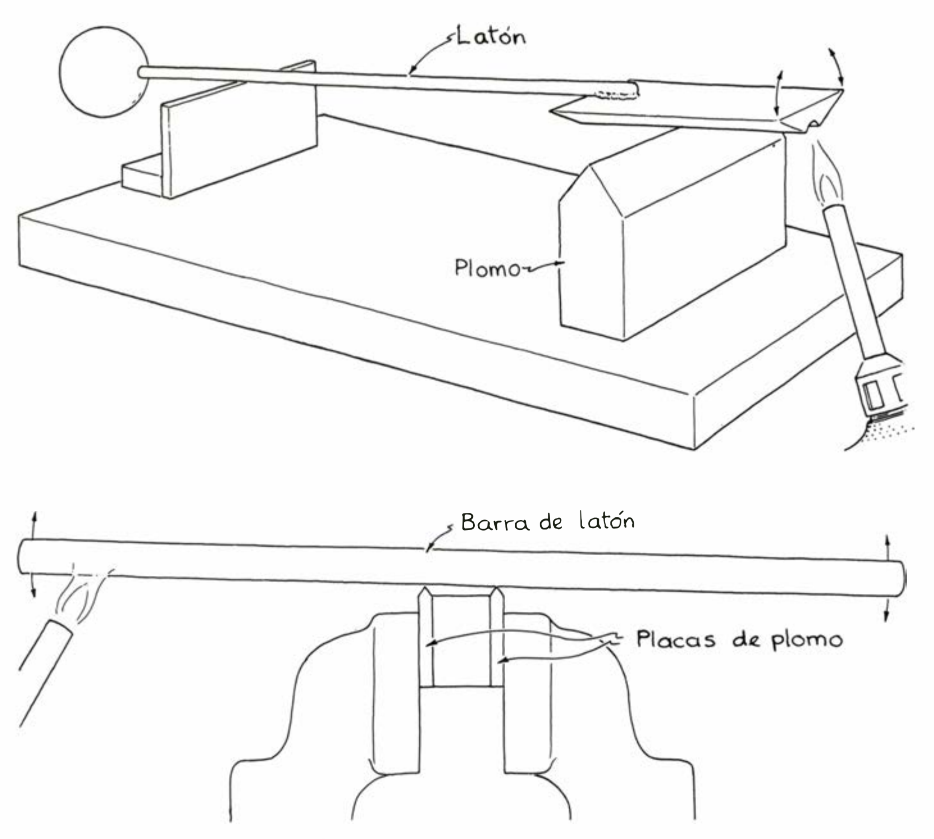
En el dispositivo de Trevelyan la varilla era de cobre, pero actualmente es más frecuente el latón. Acaso haya otros metales que puedan emplearse en lugar del cobre, latón o plomo, pero no todos sirven para producir oscilaciones sostenidas. Parece ser que fue Sir John Leslie el primero en ofrecer una explicación de las oscilaciones y el tono a ellas asociado poco después de que Trevelyan presentara su balancín. Investigaciones recientes la han depurado pero no la han sustituido.

Centrémonos en el caso del trozo de latón que descansa sobre un bloque de plomo. Las zonas de contacto son los dos bordes afilados del trozo de latón. Cuando una llama calienta la varilla y el trozo de latón, al plomo le llega calor a través de los dos bordes. La velocidad de transferencia por un borde depende parcialmente del peso que descansa sobre él; a mayor peso, mejor contacto y mayor transmisión de calor. Inicialmente, uno de los bordes transmitirá mejor el calor y, merced a la dilatación térmica, en el plomo subyacente se formará un diminuto montículo. Si esta dilatación sobreviene con rapidez suficiente, proyectará el filo hacia arriba.

Ocurren, luego, dos cosas. Como uno de los filos ha sido despedido hacia arriba, la totalidad del peso del trozo de latón se trasladará al otro filo; el calor se transmite, así, al plomo en contacto con éste. Por otro lado, se enfriará y contraerá el abultamiento bajo el primer filo, mientras que se forma otro bajo el segundo. De este modo, oscila el sistema a la vez que se repite el ciclo de formación y contracción de los bultitos.



4. Montaje para hacer que el agua oscile entre líquido y hielo



5. Balancín de Trevelyan en sus dos versiones

La cadencia del golpeteo depende parcialmente de las dimensiones del balancín. En ciertos casos, aquélla es tan elevada que se percibe un tono continuo.

¿Por qué hay algunos materiales que no sirven? Una razón es que el bloque debe ser de un material que se dilate al calentarse. Así, el plomo se dilata bien, más que el latón y el cobre. Otra es que la oscilación del balancín sólo puede mantenerse si no se dispersa apresuradamente el calor transmitido al plomo situado bajo cada filo. Pero si el calor se reparte a velocidad excesiva, esa zona del bloque no dispondrá de tiempo suficiente para dilatarse. El plomo es un conductor térmico lo bastante deficiente para que resulte muy adecuado a estos propósitos.

Una tercera razón es que el bloque debe conducir el calor con rapidez suficiente para que éste no permanezca demasiado tiempo en la zona de los filos; si no, la dilatación térmica persistiría demasiado tiempo y los bultitos menguarían con demasiada lentitud. Y los filos ya no saldrían despedidos del bloque.

Podría ser interesante invertir los papeles del latón y del plomo, disponiendo un trozo de plomo provisto de dos filos sobre un bloque de latón. Desde luego, no se puede invertir el método de calentamiento: el material a calentar sigue siendo el latón. Sería, además, divertido observar las oscilaciones de ambos sistemas bajo luz estroboscópica. El movimiento es muy leve, pero puede amplificarse mediante una aguja o una varilla delgada fija al balancín formando un ángulo recto con su eje mayor. El extremo de la aguja recorrerá una distancia mayor que el balancín, con lo que facilitará la observación del movimiento.

Hay otro oscilador térmico que, tiempo atrás, constituía una experiencia tradicional en las cátedras de física. Un trozo de tela metálica se introduce por la parte inferior de un tubo vertical hasta más o menos un cuarto de la longitud de éste. La tela metálica se dobla por los bordes para inmovilizarla dentro del tubo. Bajo éste se coloca un mechero, que caliente el tubo y la malla metálica. Si se retira el mechero, el tubo ulula algunos segundos. El tono que emite es casi puro: se trata de la frecuencia fundamental de las vibraciones acústicas en su interior.

Este sistema fue descubierto en 1859 por Pieter Leonhard Rijke, de la Universidad de Leiden, quien atribuyó el sonido a las contracciones térmicas del tubo y la malla metálica cuando éstos se enfrían, creyendo que las vibraciones que acompañan a tales contracciones excitaban la onda acústica emitida por el tubo. Rijke se equivocó: ni la malla ni el tubo vibran.

La explicación correcta la dio Lord Rayleigh en 1878. El sonido lanzado por el tubo lo originan las vibraciones forzadas del aire que contiene. El mechero encendido calienta ese aire y establece una corriente convectiva ascendente que recorre el tubo atravesando la malla metálica. En esta fase no se emite sonido alguno porque la corriente de aire es estacionaria.

Cuando se retira la fuente de calor, la corriente convectiva persiste durante un lapso, ya que el tubo y la malla están los dos calientes. Entonces, las numerosas perturbaciones que sufre el aire dentro del tubo favorecen la creación de una onda acústica estacionaria en el seno del mismo. Voy a describirles primero cómo es una onda de esta naturaleza dentro de un tubo no calentado y, luego, les mostraré de qué modo la corriente convectiva dentro del tubo de Rijke la refuerza hasta el punto de que éste ulule.

Dentro de un tubo recto puede crearse una onda acústica estacionaria mediante la interferencia de dos ondas sonoras que se propaguen en sentidos contrarios por su interior. Estas ondas deben ser iguales, salvo en su dirección de propagación. En su recorrido, atraviesan todo el tubo, se reflejan en los extremos abiertos y vuelven a recorrer el tubo. Cuando se cruzan, se establece una interferencia que obliga a las moléculas de aire a oscilar levemente en dirección paralela al eje del tubo. En virtud de la repetición de la interferencia de las ondas, las moléculas se ponen a oscilar obedeciendo a un patrón organizado. Las moléculas que se encuentran en determinados lugares de la longitud del tubo no se mueven nunca; esos lugares se llaman nodos. Por otra parte, las moléculas que ocupan otras posiciones oscilan con una amplitud máxima determinada; esos lugares se llaman vientres.

Este modelo de interferencia repetida de dos ondas progresivas recibe el nombre de onda estacionaria, porque se mantienen fijas las posiciones de los nodos y vientres dentro del tubo. Recibe el nombre de onda fundamental, o primer armónico, la onda estacionaria más sencilla que pueda formarse dentro de un tubo. Su configuración está constituida por un nodo en el centro del tubo y un vientre en cada extremo. O sea, las moléculas contiguas a los extremos oscilan paralelamente al eje del tubo con una amplitud máxima, mientras que las situadas en el centro de éste oscilan muy poco o nada.

Para excitar el primer armónico de un tubo, la frecuencia de las ondas sonoras progresivas debe ser igual a la velocidad del sonido dividida por dos

veces la longitud del tubo. Una onda estacionaria, una vez iniciada, puede aumentar de intensidad si se comunica más energía a las ondas sonoras. Se dice en tal caso que el tubo resuena a la frecuencia de la onda estacionaria fundamental, o primer armónico. Cuando se alcanza la resonancia, el sonido que salga del tubo puede ser muy fuerte.

Toda onda acústica estacionaria en el interior de un tubo establece asimismo variaciones periódicas en la presión del aire a lo largo del mismo. Consideremos el nodo central cuando se genera el primer armónico. A cada lado de éste, las moléculas oscilan paralelamente a la longitud del tubo. Durante una parte de la oscilación, las moléculas que hay a uno y otro lado se mueven hacia el nodo, aumentando la presión en él. Medio ciclo de oscilación después, las moléculas se alejan del nodo produciendo en él una baja presión. O sea, a la vez que oscilan las moléculas, la presión en el nodo crece por encima, y desciende por debajo, de la presión atmosférica reinante en el exterior: la variación de presión máxima en el interior del tubo se da precisamente en el nodo.

Hay textos de física en los que los nodos y vientres de las ondas acústicas estacionarias se definen en función de las variaciones de presión. De ese modo, los vientres de presión (o variaciones de presión altas) coinciden con los nodos de oscilación y los nodos de presión (o variaciones de presión nulas) con los vientres de oscilación. En el estudio de Rayleigh acerca del tubo de Rijke la clave se encuentra en las posiciones relativas de la malla metálica y el nodo de oscilación (o vientre de presión) central. El tubo debe estar vertical para que el calor fuerce una corriente convectiva por su interior. La malla debe estar en la mitad inferior, no demasiado cerca de la boca inferior ni del nodo de oscilación central.

Admitamos que las distintas perturbaciones que sufre el aire al enfriarse inician, de uno u otro modo, ondas progresivas débiles que generan el primer armónico del tubo. Entonces, el aire oscilará atravesando la malla metálica. Durante una parte de la oscilación, el aire deberá ascender a través de la malla en dirección al nodo de oscilación, incrementando la presión en éste. Durante la otra parte del ciclo, el aire deberá descender a través de la malla alejándose del nodo, produciendo en éste una baja presión.

Según razona Rayleigh, la oscilación del aire dentro del tubo recibe la ayuda del calor que le transmiten al aire la malla y el tubo calientes. Pero esa transmisión sólo es efectiva y útil durante una parte del ciclo. Cuando el aire asciende a través de la malla en dirección al nodo, al tubo afluye aire relativamente frío por la boca inferior. Por ello crece la diferencia de temperaturas entre el aire y la malla; ésta cede una buena cantidad de calor al aire, que pierde densidad con la temperatura. El calor cedido por la malla impulsa, pues, la masa de aire hacia arriba, en dirección al nodo, reforzando el incremento de presión y avivando la oscilación.

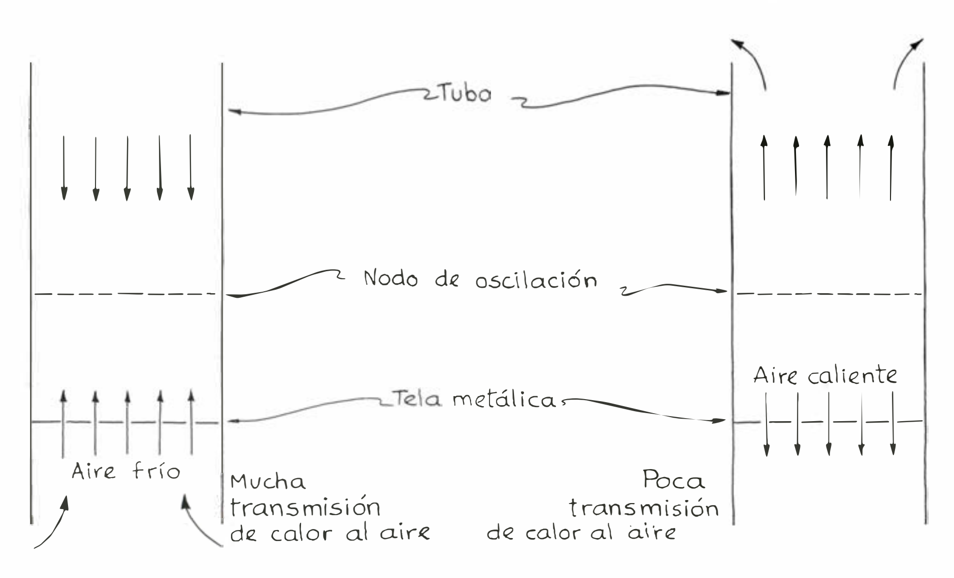
Durante la fase descendente de la oscilación del aire, el calor transferido a éste tiende a hacerlo flotar hacia arriba. Pero esta vez, el empuje de Arquímedes actúa en sentido contrario al favorable a la oscilación; aunque, como es reducido, no entorpece la vibración acústica. El aire descendente se halla casi tan caliente como la malla, pues ya había atravesado esa zona; por tanto, la transmisión de calor de la malla al aire será pequeña. La transmisión de calor sólo sirve entonces para aminorar la oscilación, pues toda adición de calor al aire propenderá a hacerlo ascender, y no a descender en sincronía con la oscilación.

La oscilación del aire en el primer armónico se ve fortalecida por la asimetría de la transmisión de calor durante las dos semioscilaciones. En efecto, la oscilación recibe un gran impulso durante medio ciclo y prácticamente ninguno durante el otro medio. Rayleigh comparó esta asimetría con una fuerza aplicada periódicamente a un péndulo oscilante. Imaginemos un péndulo al que se empuja cada vez que pasa por el punto más bajo. En una parte del balanceo, el empujón supondrá un aumento de energía, pues se da en el sentido del movimiento. En el siguiente paso del péndulo por el punto más bajo, el empujón se dará, empero, en sentido contrario al movimiento y ello supondrá restarle energía. En tanto se aplique siempre la misma fuerza, el péndulo no recibirá una cantidad de energía neta y el balanceo acabará mermando a causa de la pérdida de energía por rozamiento. Para reforzar el balanceo debe empujarse sólo cuando se aumente la energía de la oscilación.

Existe otra ventaja en la cesión de calor al aire cuando éste asciende hacia el centro del tubo. Durante el aumento de presión del aire situado en torno al centro del tubo, la temperatura y la presión de aquél son proporcionales. Así, cuando la malla calienta, aumenta la temperatura, acrecienta además la presión en dicha zona y, cuando el aire comienza a alejarse de ella, y por ello disminuye la presión, toda adición de calor al aire contribuirá a incrementar la presión. Por tanto, el hecho de que la transferencia sea pequeña durante esa etapa resulta beneficioso para las oscilaciones.

Cualquiera puede poner a resonar un tubo de Rijke, con tal que esté hecho de un material que no se funda al calentarse. De malla metálica puede servir perfectamente tela metálica de ventana, de la que es mejor emplear una capa doble y no una sola. El calor puede obtenerse de un mechero de laboratorio, de un fuego de cocina e incluso de una vela.

Cuando se inicie el sonido pruébese a colocar el tubo horizontal y luego devuélvase a la posición vertical. Se verá que, cuando el tubo se pone horizontal, el sonido cesa al instante, ya que la convección natural desaparece en dicha posición. Pero si el tubo vuelve a ponerse vertical antes de que la malla se haya enfriado, el sonido se reanudará.



6. Movimiento del aire en el tubo de Rijke

Libros

Biofísica, la demografía como polémica y ciencia hipocrática

Luis Cornudella, Pedro Pérez y Puchal y José Alsina

STRUCTURAL MOLECULAR BIOLOGY: METHODS AND APPLICATIONS. Dirigido por D. B. Davies, W. Saenger y S. S. Danyluk. Plenum Press, New York, 1982. Este volumen no es ningún libro de texto al uso. Se trata de una monografía perteneciente a la serie sobre ciencias de la vida que periódicamente edita la División de Asuntos Científicos de la Organización del Tratado del Atlántico Norte (OTAN). Dicha institución patrocina cinco series: la mencionada, ciencias físicas, ciencias matemáticas, ciencias sociales y ciencias aplicadas. Se recogen, aquí las aportaciones a los cursos conocidos con el nombre genérico de Institutos de Estudios Avanzados de la OTAN. La mayoría de los volúmenes, aunque no la totalidad, recopilan los programas teórico-prácticos desarrollados en los cursos monográficos o institutos avanzados bajo los auspicios de la Organización. La finalidad genérica de estos institutos es la de ofrecer los últimos logros obtenidos en campos o temas específicos a sectores especializados de la comunidad científica.

Este tipo de publicaciones son particularmente sensibles a un rápido envejecimiento, ya que abordan temas que, por su relevancia, están sometidos a ritmos acelerados de investigación y, en consecuencia, los resultados y las ideas que éstos generan cambian constantemente. A pesar de este inconveniente, poseen características muy positivas. En primer lugar, reúnen aportaciones de conjuntos significativos de calificados especialistas en distintas áreas. La posibilidad de congregar nutridos grupos de expertos en una misma ocasión es cada vez más remota debido a los efectos dispersores ocasionados por la incesante proliferación de reuniones y congresos de prestigio que se celebran anualmente. Ello entorpece la deseable coincidencia temporal de especialistas y, por ende, sólo se consiguen visiones fragmentarias de los adelantos que se producen en los distintos campos.

Un segundo aspecto a destacar es el enfoque didáctico que se imprime a estos cursos y que se mantiene en las publicaciones. En los cursos participan también postgraduados que inician estudios de especialización. Los científicos que intervienen deben esforzarse en hacer comprensibles las refinadas técnicas empleadas en los trabajos de investigación, mediante detalladas exposiciones introductorias de los fundamentos teóricos que las soportan. En los textos correspondientes a los programas desarrollados se compagina, escrupulosamente, niveles elevados de especialización con exposiciones de alto contenido didáctico. Otro elemento merecedor de especial atención es el énfasis que se confiere a las perspectivas y posibilidades que se ofrecen en cada tema tratado con el propósito de promover la innovación y ampliación de las líneas de investigación.

Al margen de la efímera actualidad inherente, ya comentada, debe mencionarse un inconveniente del que sistemáticamente adolece esta serie de la organización atlántica. Se trata del excesivo intervalo de tiempo que transcurre entre la celebración de los cursos y la aparición de las correspondientes publicaciones. Esta dilación agrava el defecto señalado. Hemos de señalar que no se produce intento alguno para paliar sus efectos, como sería la adición al final de cada capítulo de notas de actualización, insertadas inmediatamente antes de la aparición de cada volumen.

El libro que da pie a nuestro comentario es el volumen 45 de la serie sobre ciencias de la vida. Se recogen las sesiones plenarias o lecciones magistrales del curso avanzado dedicado a la metodología empleada en biología molecular estructural, celebrado en Maratea, Italia, en mayo de 1981. La organización de este curso la compartió la Federación Europea de Sociedades de Bioquímica (FEBS); de ahí que se corresponda con el número 78 de la serie de cursos que patrocina la FEBS. Destáquese la oportunidad en la selección del tema objeto del curso: coincide con el escogido por la Real Academia de Ciencias sueca para la concesión del Nobel de química en 1982, año en que precisamente aparece el volumen que recoge las contribuciones a dicho curso. Esta circunstancia pone de relieve la importancia que se atribuye a las técnicas y métodos biofísicos; el grado de refinamiento alcanzado por los mismos y su directa incidencia en el avance del conocimiento sobre la naturaleza de los mecanismos que mediatizan y regulan a los seres vivos.

El volumen contiene tres secciones y un apéndice. A modo de preámbulo, la primera introduce al lector en los principios que rigen la organización biológica y su aplicación a las estructuras moleculares. La segunda parte se dedica por entero al desarrollo conceptual de las grandes técnicas físicas, entendiendo por tales aquellas cuyo tangible poder resolutivo nos desvela los detalles estructurales más finos de la organización molecular. Característica común de todos los capítulos que integran esta sección es la glosa del nivel de refinamiento conseguido por las distintas técnicas que sustentan la elaborada experimentación que se realiza en el campo estructural de la biología molecular.

La relación metodológica es exhaustiva. Uno de los capítulos iniciales aborda el desarrollo de la cristalografía alcanzado en el estudio de proteínas globulares. Sigue un conjunto de artículos cuya finalidad es explicar el fundamento teórico de la difracción de rayos X por fibras y la dispersión de neutrones a bajo ángulo, así como los detalles de su aplicación práctica. Un segundo grupo de contribuciones recrean el alcance de la resonancia magnética nuclear de alta resolución y su aplicación específica al reconocimiento celular, al estudio de los ácidos nucleicos y las conformaciones adoptadas por las proteínas. No se olvida mencionar las aportaciones realizadas por la espectroscopía Raman de resonancia en el examen de las macromoléculas biológicas. El máximo exponente tecnológico lo ofrece el capítulo reservado a la visualización de micrografías electrónicas de estructuras moleculares por reconstrucción de imagen. Se cierra esta parte del libro con cuatro capítulos que glosan: *a*) los métodos ópticos aplicados a los sistemas biomoleculares (dicroísmo circular, dispersión óptica rotatoria, luminiscencia polarizada circular y dispersión diferencial de intensidad circular); *b*) las técnicas calorimétricas en el estudio de la estructura y organización biomoleculares; *c*) los ensayos para el estudio de cinéticas aceleradas en sistemas biológicos; y *d*), las funciones energéticas potenciales en las

interacciones inter e intramoleculares. La tercera sección, la más breve, presenta una estrategia opuesta a la empleada en la precedente. Para resaltar la versatilidad de las metodologías examinadas, atenuada quizá por la exigencia de las disquisiciones teóricas impuesta en las exposiciones de la primera parte, se plantean cuatro problemas específicos, detallándose a continuación los pormenores de los tratamientos técnicos empleados para avanzar en su resolución. Esta sección contiene la aportación del único representante español que intervino en el curso. Su tema versa sobre el estudio por difracción de rayos X de las interacciones ADN-proteínas, en complejos reconstituidos y en estructuras nativas.

Finalmente, el apéndice recoge las restantes actividades realizadas durante el curso; a saber: *1*) los seminarios científicos celebrados con sus títulos y presentadores; *2*) trabajos divulgados en forma de paneles con sus autores; *3*) relación completa de los científicos participantes en el curso con sus instituciones de procedencia.

Es evidente que las publicaciones que recogen las series de los Institutos Avanzados de la OTAN no están pensados para un público no especializado, ni tan sólo para universitarios de primer ciclo. Van dirigidos a sectores específicos de la comunidad científica, y, por extensión, a graduados de tercer ciclo en el umbral de su especialización. El volumen concreto objeto de este comentario será, sin duda, de gran valor para los investigadores en biología estructural de nivel molecular; ofrece una completa recopilación y actualización de toda la metodología y la información más reciente. La carencia de uniformidad expositiva, habitual en los textos confeccionados por muchos autores, es imperceptible en el presente caso. Ello es debido, con toda probabilidad, al esquema estricto exigido por los editores a los distintos firmantes para la redacción de su artículos, notándose el efecto normalizador que esta imposición conlleva. Asimismo, es de resaltar el acierto demostrado en la confección de dicho esquema al insistir en los fundamentos teóricos y conceptuales por una parte y la aplicación práctica de la metodología por la otra, con objeto de hacer esta última inteligible. Debe reconocerse el notable equilibrio conseguido en este empeño: el desarrollo global de ambos enfoques resulta perfectamente equiparable. (L. C.)

DIEZ ERRORES SOBRE LA POBLACIÓN ESPAÑOLA, por Amando de Miguel. Madrid, Tecnos, 1982. "Mon cher, fais un peu de statistique, science assez utile quan on n'en abuse pas" (H. Balzac: *Les illusions perdues*). La fértil minerva de Amando de Miguel ha dado nuevamente uno de sus frutos, esta vez referido al estudio de la natalidad española. Como el tono polémico es el que mejor cuadra a su incisivo estilo literario, este librito está presentado, en composición, tipografía y hasta en el título, de manera que los datos y la argumentación se desarrollen en ataque a errores reales o supuestos. Para ello se distribuyen en diez capítulos, cada uno de los cuales está encabezado con un texto en cursiva, recuadrado, y junto al que campea una especie de matasellos, que previamente descalifica la opinión en él expresada, porque en negrillas capitales de gran tamaño se lee la palabra **ERROR**. Y luego el capítulo parece dedicado a demostrar ese yerro, cosa que consigue en pocos párrafos y, venturosamente, el resto del capítulo se destina a otras cuestiones, que suelen tener mayor interés. De esos diez errores, algunos son ciertos y están extendidos entre la gente común, como son el de que nacen más niñas que niños, o el de que la fecundidad de las zonas cálidas es superior a la de las zonas frías en España; errores que destruye fácilmente, aunque en el primero añade el error suplementario de que mueren más niñas que niños, opinión que no se sabe quién puede sustentarla y que aparentemente se utiliza más que nada para atribuirla a prejuicios machistas y *debelare superbos*. Los demás errores que encabezan sendos capítulos presuponen, para que se pueda incurrir en ellos, estudios o conocimientos demográficos, pero claro es que esos conocimientos impiden que muchos de ellos puedan mantenerse. De modo que, dado el tono de desafío que se emplea a menudo, más bien sugieren empeños en atribuir a la gente prejuicios, bien religiosos, bien machistas, bien éticos o políticos, que el autor se complace en deshacer, aunque frecuentemente no haga otra cosa que "desentrañar obviedades", que diría Torrente Ballester, o alancear moros muertos.

Un ejemplo típico de esta actitud es el reproche a toro pasado de la falta de acierto de una previsión, como si de acertijos se tratara, cuando lo único reprochable es extrapolar desacertadamente una tendencia; así, cuando critica sendos textos de un científico en 1920 y de un historiador en 1960, que señalan los peligros para la sociedad de la baja natalidad, simplemente porque escriben en los umbrales de los decenios caracterizados por el crecimiento demográfico, olvidando deliberadamente (pues no es por ignorancia) que esos umbrales son también el cierre de los anteriores decenios en los que se producen bajas de fecundidad, notables en relación con la mortalidad vigente en los respectivos tiempos. En cambio, se enorgullece de sus propios aciertos, a veces casuales, y cuenta cosas tan divertidas (pág. 54) como su vaticinio a finales de 1980, ante sus alumnos y sólo con los datos del rápido descenso de la natalidad, de que pronto tendríamos una cruel sequía, un golpe de Estado y una epidemia; y, en efecto, durante 1981 se produjeron las tres cosas: la intentona del 23-F, la intoxicación de la "colza" y la sequía. Realmente, no se puede adivinar más a partir de una simple curva de natalidad bruta y nada podría admirar más a un meteorólogo que comprobar los efectos de la cantidad de nacimientos en el comportamiento del frente polar, ni a Malthus ver a sus "frenos positivos" actuar de tan contradictoria manera. También De Miguel hace correctas previsiones de un futuro inmediato partiendo de la tendencia observada en el quinquenio 1976-80, que tiene visos de continuar, pero que bien podría detenerse y hasta invertirse, quizás en los próximos diez años.

Otro ejemplo de argumentación es el del error que encabeza el primer capítulo y que se compone de dos partes: una, que en los países europeos está bajando la natalidad; otra, que, en España, las familias numerosas son la norma estadística. Pues bien, en el mismo enunciado de esos errores y en su interrelación está la mixtificación o, si se quiere, la "descripción del maniqueo" que se propone vapulear. Nadie medianamente enterado cree que en una serie de países europeos noroccidentales esté bajando la natalidad, sino que ha bajado tanto que difícilmente puede continuar descendiendo. Y ¿quién cree en España, incluso entre la gente común, que la familia numerosa sea la normal? ¿Quién no sabe que hoy (y ese hoy se remonta a dos o tres decenios) es raro el matrimonio joven que se avenga a tener más de dos o tres hijos, que pocos llegan a los cuatro y que los que tienen más o son viejos o se trata de excepciones? De Miguel está en lo cierto cuando apunta el continuo descenso de la fecundidad española que, evidentemente, está llegando, por cierto, al límite peligroso en que las generaciones empieza a no ser sustituidas enteramente por las siguientes, pero no tiene derecho a afirmar tan categóricamente, como hace en este libro y repite en la prensa y por la radio, que en las dos Alemanias y en Estados Unidos,

cuyas estadísticas trae a cuento, o en los demás países europeos la tendencia de la natalidad sea ascendente, porque las cifras que maneja son las de 1977 a 1980 y sabido es que, aunque las técnicas estadísticas permiten aproximarse mucho al calcular los índices año por año, sin embargo, los de los años intercalados entre dos censos sucesivos (que suelen ser decenales o quinquenales) carecen de precisión absoluta, al tener que estimarse en cada año la población media por interpolación o, lo que es menos preciso, por extrapolación de la tendencia del período anterior, como es el caso que comentamos. En resolución, que el hecho de que el índice de 1980 sea superior en unas décimas a los de los años inmediatamente anteriores, puede no tener ningún significado; aparte de que una ligera oscilación en el número de nacimientos tampoco permite hablar de tendencia, porque tales fluctuaciones son completamente naturales. Máxime, cuando en Alemania federal se ha alcanzado una tasa del 9,5 por 1000 y una reproducción neta del 0,6, lo que augura que, de seguir así, la generación siguiente menguaría en un 40 por 100 respecto de la de sus progenitores. Eso revela que el *Zweikindersystem* o número generalmente deseado de dos hijos, que imperó en los años veinte en Alemania y en otros países y que, dada la mortalidad de entonces, puso la reproducción neta por debajo de uno, ha sido sustituido por el *Einzelkindsystem* o ideal del hijo único, que ya, ni con la bajísima mortalidad actual, hace posible la renovación de las generaciones; por consiguiente, es previsible una próxima reacción que evite el lento suicidio del pueblo alemán y el de otros europeos en análogas circunstancias. ¿Tiene, pues, algún significado que en un año la bajísima natalidad bruta haya subido el 0,6 por 1000 e incluso el 1 por 1000, salvo que esa esperada reacción hubiera empezado a manifestarse? Pues sí, el demostrar que el divorcio favorece la natalidad y que la "píldora" y el aborto no la disminuye (pág. 21). A uno le parece mentira que De Miguel regañe en el prólogo a los demás por la *utilización ideológica* de los datos y el "no ver lo que no se quiere ver" y, en cambio, sea incapaz de reparar en la viga de su propio ojo.

En el capítulo tercero, se empeña en combatir el presunto error de los teóricos de la *transición demográfica* sobre que las características vitales de los países subdesarrollados de hoy son las mismas que las que tuvieron los actuales países desarrollados. Para ello, compara los índices españoles de natalidad y fecundidad de finales del siglo XIX con los actuales de varios países centroamericanos. Pero toda su argumentación ignora tres hechos fundamentales: *a*) que a finales del siglo XIX la natalidad española había ya comenzado a bajar, según revelan los estudios de Bacci sobre la natalidad del año 1768; *b*) que en Europa, y también en España, la edad media de matrimonio era más tardía en general que la de los actuales países en desarrollo; y, *c*) en todos los países americanos y africanos se ha observado un aumento de la natalidad como resultado de las eficaces y rápidas campañas sanitarias que, junto con la mortalidad general y perinatal, han hecho disminuir en picado las esterilidades y los embarazos frustrados.

En el capítulo 5.°, el más extenso y de los más logrados del libro, se inquiere las causas del descenso de la fecundidad, bien que amparado en uno de esos errores gratuitos que él levanta, pero que poco tiene que ver con lo que en el texto se discute. Allí se examinan como posibles causas el desarrollo económico, la profesión, la urbanización, el nivel de instrucción, la edad, la religión, el trabajo femenino, con conclusiones interesantes, muchas ya sabidas, y dedica también muchas páginas a la en otro tiempo batallona cuestión de las clases sociales y su efecto sobre la natalidad. Es harto conocido, sin embargo, tras las muchas comprobaciones contradictorias según las épocas y los lugares, que las prácticas anticonceptivas, como sucede con tantas normas de cultura o de conducta social, penetran por las clases más ilustradas, que solían ser las más elevadas y, con un efecto *trickle down*, se van filtrando paulatinamente hacia la base hasta llegar a generalizarse por toda la "pirámide social" hasta que, una vez alcanzado este estado, el tramo siguiente es el de producirse una mayor fecundidad en las clases con más recursos para sostener la prole, lo cual constituye, justamente, la situación inversa a la del comienzo del proceso. Claro es que el estadio de esa evolución es distinto en unas comunidades y en otras, en las grandes ciudades, en las pequeñas o en el medio rural y, además, este esquema general puede estar alterado por circunstancias tales como el grado de instrucción y el trabajo femenino (que son determinantes) o la religiosidad, aunque esta última sea cada vez menos influyente, como el autor se complace en señalar.

Pero el libro presenta capítulos decididamente felices. Así, el 9.°, en el que, mediante la comparación de la natalidad, fecundidad general y fecundidad legítima con los porcentajes de mujeres casadas en determinados grupos de edad fértil y de matrimonios celebrados respecto al número de mujeres solteras de la misma edad, distinguiendo las edades más fértiles y las más frecundas, logra determinar el quinquenio 1965-70 como aquel en que muchas mujeres jóvenes deciden retrasar considerablemente su primer embarazo; y así mismo, que el ascenso de la natalidad en 1960 se debe más a la nupcialidad e incremento de primogénitos que a la fecundidad o descendencia final.

Así también, el capítulo 7.°, donde se impugna las ideas increíblemente racistas de Vandellós respecto a catalanes frente a murcianos y andaluces orientales. Es curioso la impregnación de las ideas coetáneas del nazismo, incluso entre quienes no participaban de sus concepciones totalitarias. Pero, en fin, cada uno es hijo de su tiempo y, en los años treinta de recesión económica y aparente descrédito del demoliberalismo, las ideas totalitarias de alemanes, italianos y soviéticos ejercían una influencia innegable hasta en sus mismos adversarios. Sin embargo, al impugnar en un todo las ideas de Vandellós, De Miguel no acaba de percatarse de que precisamente la posterior y reciente recuperación de la natalidad en Cataluña viene a confirmar los temores de Vandellós, ya que se debe más que nada a la inmigración, que amenaza sumergir la cultura específicamente catalana, empezando por su lengua, y de ahí el empeño actual de catalanizar a los hijos de los inmigrantes desde la misma infancia y aún preescolaridad.

En resumen, un libro muy en la línea de Amando De Miguel, polémico, audaz, divertido a veces, hiriente otras, ilustrador a menudo, aburrido nunca. (P. P. P.)

THE HIPPOCRATIC TREATISES "ON GENERATION", "ON THE NATURE OF CHILD", "DISEASES, IV", por Iain M. Lonie. Ars Medica; W. de Gruyter, Berlin; 1981, 406 pp. De nada estamos más necesitados los estudiosos de la medicina antigua que de ediciones de los distintos tratados, con un nuevo texto que venga a sustituir el benemérito, aunque deficiente, de Littré, y un buen comentario. Si lleva, además, una traducción, mejor aún. En el campo de la medicina hipocrática, sin ser muchos, son ya algunos los tratados objeto de edición, traducción y comentario. Desde hace algún tiempo, parece que los filólogos e historiadores de la medicina se han dado cuenta de esa necesidad y comienzan a ocuparse de colmar tan importante laguna: Unger, en 1923, nos ofreció un estudio de este tipo so-

bre el tratado *De corde*; casi diez años después, K. Deichgräber publicaba, en colaboración con Schwyzer, el *De carnibus*. Siguieron luego, sin intervalos tan largos entre sí, los de Festugière (*De vetere medicina,* 1948), Grensemann (*De morbo sacro, De octimensi partu*, 1968), Potter (*De morbis*, III, 1970), Diller (*Ueber die Umwelt*, Berlín, 1970), Wittern (*De morbis I*, Hildesheim, 1974), Jouanna (*De natura hominis*, Berlín, 1975)... Añadiremos que nuestra edición de los *Tratados médicos* de Hipócrates (Barcelona, Alpha) tiene ya su tercer tomo a punto de aparecer (este tercero contiene *Epidemiae I* y *III* y *De natura hominis*).

Estamos, pues, en cierto modo, de enhorabuena: porque ahora acaba de aparecer una traducción, con introducción y un amplísimo comentario, de uno de los tratados hipocráticos más importantes de la colección: *De genitura, De natura pueri, De morbis IV*, que ha sido encargado al prestigioso filólogo e historiador de la medicina Iain Lonie, bien conocido por sus trabajos sobre los tratados cnidios del *Corpus*, entre otros.

No se trata de una edición, hay que decirlo inmediatamente. Lonie parte del texto de Littré. Quizá no sea una buena decisión, habida cuenta del poco, escasísimo valor que la edición de este erudito, tan benemérito, por otra parte, tiene para el estudioso un poco exigente. Posiblemente haya pesado, a la hora de tomar la decisión, el hecho de que apareciera, hace unos pocos años, dentro de la colección Budé y a cargo de R. Joly, una buena edición, con traducción y notas (en esta misma colección y por obra del propio Joly han salido ya cuatro volúmenes, que vienen a engrosar la lista de ediciones parciales que antes hemos ofrecido). Sin embargo, posiblemente haya pesado el factor económico: los textos griegos encarecen muchísimo los costes. Y digo esto porque Lonie no se remite a la edición de Joly, sino a Littré, aunque, obligado es decirlo, aportando numerosas correcciones, que señala y discute en el amplísimo comentario.

La estructura del libro es la siguiente. Se abre con la versión, con numeración seguida, de los tres tratados. Ya desde Littré es ésa la práctica común, por creerse que se trata de tres tratados que forman una unidad, no ya de autor, sino de tema. Con poquísimas excepciones (Kahlenberg, por ejemplo, *Hermes*, 90-1955, 252 ss., opina que *De morbis IV* es una adición posterior; Fredriech creía en 1899 lo contrario: que este opúsculo era anterior; Joly sostiene que *De morbis* se agregó con posterioridad al grupo formado por *Genitura* y *Natura pueri*). Antecede a la traducción una amplísima bibliografía (29 páginas de apretado texto). Tras la versión, una introducción que aborda los siguientes puntos: unidad del tratado e identidad del autor, intención de la obra, relación con los demás tratados hipocráticos, la teoría de los humores, relación con las doctrinas presocráticas, fecha del tratado, método científico del autor, y, finalmente, un largo apéndice, donde se estudia el sistema vascular del tratado comparado con el que se puede reconstruir a partir de los restantes tratados hipocráticos. Se nos ofrece un amplísimo comentario (de hecho, dos terceras partes de la obra) y se cierra con tres índices: un índice general, otro de términos griegos que aparecen en el comentario y un *index locorum*. Una estructura muy habitual en este tipo de libros, en especial los que publica esta colección.

Ante la imposibilidad de hacer una crítica exhaustiva del contenido del libro, nos limitaremos a unos cuantos aspectos que creemos imprescindibles para darse cuenta de la importancia y riqueza de esta obra.

Comencemos por el texto. Hemos señalado ya que Lonie parte del texto de Littré. Pero cuando se aparta de él, o ve razones suficientes para mostrar cierto escepticismo de Littré, una nota

del comentario lo justifica. De este modo podemos señalar una serie de puntos donde Lonie no sigue a Littré. Sin ánimo de agotar todos los pasajes donde esto ocurre, señalaré sólo algunos. En *Genitura* se aparta o pone objeciones a la lectura de Littré en: I, 1, que considera corrupto (+ τείνουσι ... ἀποχωρέειν +); IV, 1, donde elimina [αὐτὴν], entendiendo el verbo μεθῇ como *eyacular* (*until he ejaculates*; Joly sigue aquí la interpretación de Littré: *jusqu'à ce que l'home la lache*); más correcciones en IV, 3; VIII, 1, IX, 1, XI, 1. En *De natura pueri* anotaremos los siguientes pasajes donde Lonie discrepa: XII, 1, donde ἐν θερμῷ ἐοῦσα cree el editor que debe excluirse; XII, 6, donde acepta la corrección de Joly de τὸ en τι; XIII, 1, donde lee οἰκείης; XIII, 2, XV, 4, XVII, 2, XVIII, 3, etcétera. En *De morbis IV* hay discrepancias con Littré, entre oteos, en XXXIX, 1, donde se pone en duda la lectura de los manuscritos; en XXXIX, 5 (en que defiende que la corrección de Joly es atractiva), y en XLII, 2, donde se adopta la conjetura de Ermerins.

Sobre la traducción poco es lo que tengo que decir: una lectura atenta permite darse cuenta de que es ajustada y que, en algunas ocasiones, mejora las hasta ahora existentes.

Más importante me parece el comentario, que forma un todo con la introducción. Diría que entre uno y otro no queda punto por tocar. El método empleado por el autor consiste en hacer una introducción parcial antes de cada uno de los grandes apartados: sigue, luego, el comentario pormenorizado. Se vuelcan, aquí, todas las observaciones necesarias para la intelección, gramatical y médica, del texto. En algunos casos, si ello es necesario, se establece una comparación con otros tratados hipocráticos, no sólo para explicar el aspecto terminológico (tan importante en estos casos) sino también de doctrina. No falta la polémica, siempre en un tono comedido, lo que aclara mucho los puntos oscuros –que los hay, y, yo diría, sigue habiéndolos después de Lonie– pero que resultan ya más diáfanos.

En resumen: el camino que ha tomado la investigación hipocrática prosigue con Lonie. Tenemos ahora un nuevo instrumento de trabajo para conocer mejor la lengua, las doctrinas y los métodos de Hipócrates. Ya hemos hecho alusión a la necesidad que tenemos de libros como éste. Sólo nos falta, ahora, junto a la continuación de esos tratados, que aparezca, por fin, el tan esperado *Léxico Hipocrático*, que habrá de ser un instrumento imprescindible de trabajo. (J. A.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

FLUIDEZ DE LAS MEMBRANAS CELULARES

PHASE TRANSITIONS AND FLUIDITY CHARACTERISTICS OF LIPIDS AND CELL MEMBRANES. D. Chapman en *Quarterly Reviews of Biophysics*, vol. 8, n.° 2, págs. 185-235; 1975.

INTRINSIC PROTEIN-LIPID INTERACTIONS. PHYSICAL AND BIOCHEMICAL EVIDENCE (Review Letter). D. Chapman, J. C. Gómez-Fernández y F. M. Goñi en *FEBS Letters*, vol. 98, n.° 2, páginas 211-223; 1979.

LIPID CONFORMATION IN MODEL MEMBRANES AND BIOLOGICAL MEMBRANES. J. Seelig y A. Seelig en *Quarterly Reviews of Biophysics*, vol. 13, número 1, págs. 19-61; 1980.

THE INTERACTION OF INTRINSIC PROTEINS AND LIPIDS IN BIOMEMBRANES. D. Chapman, J. C. Gómez-Fernández y F. M. Goñi en *Trends in Biochemical Sciences* vol. 7, n.° 2; 1982.

TEORIA RETICULAR DEL CONFINAMIENTO DE LOS QUARKS

EL MODELO EN BOLSA DEL CONFINAMIENTO DE LOS QUARKS. Kenneth A. Johnson en *Investigación y Ciencia*, n.° 36, págs. 76-86; 1979.

TEORÍAS GAUGE DE LAS FUERZAS ENTRE PARTÍCULAS ELEMENTALES. Gerard 't Hooft en *Investigación y Ciencia*, número 47, págs. 58-77; agosto, 1980.

LATTICE GAUGE THEORIES AND MONTE CARLO SIMULATIONS. Claudio Rebbi en *Non-Perturbative Aspects of Quantum Field Theory*. World Scientific Publishing Co;

VACUNAS SINTETICAS

CHEMICAL SYNTHESIS OF A POLYPEPTIDE PREDICTED FROM NUCLEOTIDE SEQUENCE ALLOWS DETECTION OF A NEW RETROVIRAL GENE PRODUCT. J. G. Sutcliffe, T. M. Shinnick, N. Green, F.-T. Lie, H. L. Niman y R. A. Lerner en *Nature*, vol. 287, n.° 5785, páginas 801-805; 30 de octubre de 1980.

ANTIBODIES TO CHEMICALLY SYNTHESIZED PEPTIDES PREDICTED FROM DNA SEQUENCES AS PROBES OF GENE EXPRESSION. Richard A. Lerner, J. Gregor Sutcliffe y Thomas M. Shinnick en *Cell*, vol. 23, n.° 2; febrero, 1981.

TAPPING THE IMMUNOLOGICAL REPERTOIRE TO PRODUCE ANTIBODIES OF PREDETERMINED SPECIFICITY. Richard A. Lerner en *Nature*, vol. 229, n.° 5884, págs. 592-596; 1982.

PROCESOS VISUALES OCULTOS

CYCLOPEAN STIMULATION CAN INFLUENCE SENSATIONS OF SELF-MOTION IN NORMAL AND STEREOBLIND SUBJECTS. Jeremy M. Wolfe y Richard Held en *Perception and Psychophysics*, volumen 28, n.° 2, págs. 139-142; 1980.

INTEROCULAR TRANSFER OF VISUAL AFTEREFFECTS. R. Blake, R. Overton y S. Lema-Stern en *Journal of Experimental Psychology: HPP*, vol. 7, págs. 367-381; 1981.

IS ACCOMMODATION COLORBLIND? FOCUSING CHROMATIC CONTOURS. Jeremy M. Wolfe y D. Alfred Owens en *Perception*, vol. 10, págs. 53-62; 1981.

A PURELY BINOCULAR MECHANISM IN HUMAN VISION. Jeremy M. Wolfe y Richard Held en *Vision Research*, vol. 21, n.° 12, págs. 1755-1759; 1981.

LA CORONA SOLAR

THE SOLAR CORONA. Jay M. Pasachoff en *Scientific American*, vol. 229, número 4, págs. 68-79; octubre, 1973.

A NEW SUN: THE SOLAR RESULTS FROM SKYLAB. John A. Eddy. Scientific and Technical Information Office, National Aeronautics and Space Administration, U.S. Government Printing Office, 1979.

COMPUTADOR OPTICO

OPTICAL BISTABLE DEVICES: THE BASIC COMPONENTS OF ALL-OPTICAL SYSTEMS? H. M. Gibbs, S. L. McCall y T. N. C. Venkatesan en *Optical Engineering*, vol. 19, n.° 4, páginas 463-468; julio-agosto, 1980.

OPTICAL BISTABILITY. Dirigido por Charles M. Bowden, M. Ciftan y H. R. Robl. Plenum Press, 1981.

DYNAMIC NON-LINEAR OPTICAL PROCESSES IN SEMICONDUCTORS. Allan Miller, David A. B. Miller y S. Desmond Smith en *Advances in Physics*, volumen 30, n.° 6, págs. 697-800; noviembre-diciembre, 1981.

BISTABLE OPTICAL DEVICES: PHYSICS AND OPERATING CHARACTERISTICS. David A. B. Miller en *Laser Focus with Fiberoptic Technology*, vol. 18, n.° 4, págs. 79-84; abril, 1982.

RANAS VENENOSAS

MORPHOLOGY OF THE GRANULAR SECRETORY GLANDS IN SKIN OF POISON-DART FROGS (DENDROBATIDAE). Maria Neuwirth, John W. Daly, Charles W. Myers y Lois W. Tice en *Tissue & Cell*, vol. 11, n.° 4, págs. 755-771; 1979.

TAXONOMY AND ECOLOGY OF *DENDROBATES BOMBETES*, A NEW ANDEAN POISON FROG WITH NEW SKIN TOXINS. Charles W. Myers y John W. Daly en *American Museum Novitates*, número 2692, págs. 1-23; 29 de enero de 1980.

THE PHARMACOLOGY OF PUMILIOTOXIN-B, 1: INTERACTION WITH CALCIUM SITES IN THE SARCOPLASMIC RETICULUM OF SKELETAL MUSCLE. E. X. Albuquerque, J. E Warnick, M. A. Maleque, F. C. Kauffman, R. Tamburini, Y. Nimit y J. W. Daly en *Molecular Pharmacology*, vol. 19, n.° 3, páginas 411-424; mayo, 1981.

BOVEDAS DE ABANICO

MEDIEAL STRUCTURE: THE GOTHIC VAULT. James J. Acland. University of Toronto Press, 1972.

FAN VAULTING: A STUDY OF FORM, TECHNOLOGY AND MEANING. Walter C. Leedy, Jr. Arts and Architecture Press, Santa Monica, 1980.

EXPERIMENTS IN GOTHIC STRUCTURE. Robert Mark. The MIT Press, 1982.

TEMAS METAMAGICOS

THE LITTLE LISPER. Daniel P. Friedman. Science Research Associates, Inc., 1974.

LET'S TALK LISP. L. S. Siklóssy. Prentice-Hall, Inc., 1976.

LISP. Patrick Henry Winston y Berthold K. Horn. Addison-Wesley Publishing Co., Inc., 1981.

TALLER Y LABORATORIO

THEORY OF RESONATORS. John William Strutt, Baron Rayleigh en *The Theory of Sound*. Dover Publications, Inc., 1945.

OBSERVATION OF OSCILLATORY ICE STATES IN A SIMPLE CONVECTION EXPERIMENT. Pierre Welander en *Journal of Geophysical Research*, vol. 82, número 18, págs. 2591-2592; 20 de junio de 1977.